Comparative gene expression to study the developmental basis of organ diversification.

Dissertation
for the award of the degree
Doctor rerum naturalium (Dr.rer.nat.)

Division of Mathematics and Natural Sciences
of the Georg-August-Universität Göttingen
within the doctoral program *Genes and Development*
of the Georg-August University School of Science (GAUSS)

Submitted by

Elisa Buchberger
from Vils, Tirol, Austria

Göttingen, July 2019

Thesis Advisory Committee

Dr. Nico Posnien (Supervisor)

Dep. of Developmental Biology, Johann-Friedrich-Blumenbach-Institute of Zoology and Anthropology, Georg-August-University Göttingen

Prof. Daniel J. Jackson

Dep. of Geobiology, Geoscience Centre, Georg-August-University Göttingen

Prof. Steven A. Johnsen

Dep. of Gastroenterology and Hepatology; Mayo Clinic – Rochester, Minnesota

Members of the Examination Board

First Reviewer: Dr. Nico Posnien

Dep. of Developmental Biology, Johann-Friedrich-Blumenbach-Institute of Zoology and Anthropology, Georg-August-University Göttingen

Second Reviewer: Prof. Daniel J. Jackson

Dep. of Geobiology, Geoscience Centre, Georg-August-University Göttingen

Extended Examination Board

Prof. Steven A. Johnsen

Dep. of Gastroenterology and Hepatology; Mayo Clinic – Rochester, Minnesota

Prof. Christoph Bleidorn

Dep. of Animal Evolution and Biodiversity, Johann-Friedrich-Blumenbach-Institute of Zoology and Anthropology, Georg-August-University Göttingen

Prof. Argyris Papantonis

Institute for Pathology, University Medical Center Göttingen

Dr. Gerd Vorbrüggen

RG Molecular Cell Dynamics, Max Planck Institute for Biophysical Chemistry

Date of oral examination: September 3rd, 2019

Declaration

I herewith declare, that I prepared the Dissertation '*Comparative gene expression to study the developmental basis of organ diversification*' on my own and with no other sources and aids than quoted.

_____ Göttingen, July 23rd, 2019

Elisa Buchberger

Prudens interrogatio quasi dimidium sapientiae.

(Francis Bacon)

Acknowledgements

First and foremost, I would like to thank Dr. Nico Posnien for giving me the opportunity to work on this exciting project, but especially for his optimism and trust in me and my work. I would like to thank you for always taking my opinion serious, and supporting me in every way possible, be it by scientific discussions, encouraging me to attend courses and conferences or giving me the freedom to decide in which direction the project is going. I consider myself extremely lucky to be a member of the Posnien Lab.

I thank the members of my Thesis Committee, Prof. Daniel Jackson and Prof. Stephen Johnsen. Your interest in my work and helpful discussions made the TAC meetings something I was always looking forward to. I would also like to thank Prof. Christoph Bleidorn, Prof. Argyris Papantonis and Dr. Gerd Vorbrüggen for agreeing to serve in my Extended Examination Board.

I would like to thank Prof. Ernst Wimmer for hosting me in his department for more than four years. Him, Prof. Gregor Bucher, Prof. Sigrid Hoyer-Fender, Dr. Gerd Vorbrüggen and Dr. Ufuk Günesdogan I would like to thank for the relaxed atmosphere in the department and their scientific input and advice.

A special thanks goes to Max Farnworth, for countless coffee and fruit breaks, for many deep discussions and silly jokes (and GIFs) and for always being there when needed. The PhD-journey, especially the last months, would have been so much harder and less fun without you as a friend.

I am extremely grateful to Dr. Micael Reis. Without your constant scientific and non-scientific advice this work wouldn't have turned out the way it has. Every PhD student can only wish for having such a dedicated and skilled post-doc and friend by his or her side. I learned a lot from you!

I thank Amel Chtioui, Ting-Hsuan Lu and Gordon Wiegleb for being such joyful lab mates. I would like to thank my students, especially Anıl Bilen, Sanem Ayaz, Cristina Matas de las Heras and Armin Nikšić. The times when we worked together on the project were the times when it made the biggest progress.

I would like to thank all the people in the Department of Developmental Biology, including PhD colleagues, post-docs and technicians who made working there such a fun experience. I am

especially grateful to Felix Quade, Beate Preitz and Marita Büscher for always helping when help was needed. I highly appreciate the constant support from Merle Eggers, Birgit Rossi and Bettina Hucke. Thanks to Hassan Mutasim Mohammed Ahmed for our chats not only during the weekends, Bibi Atika for always asking how I am and so many others in the lab whom I cannot all mention here. I would also like to thank Montserrat Torres-Oliva for all the help when I joined the lab during my Masters. And thanks to Max, Peter Kitzmann, Salim Ansari and Nico for our sports sessions.

I thank our collaborators Prof. Alistair McGregor, Prof. Fernando Casares, Dr. Montserrat Torres-Oliva and Dr. Isabel Almudi, for sharing their data and their valuable scientific input. I thank Dr. Barbora Konopová for our fruitful collaboration on the *Schistocerca* project. Also, this work would not have been possible without the bioinformatics community and all the researchers, that publish their codes and programs open access.

I am very grateful to the team of the GGNB office. Their ongoing support with all organizational things makes everything much easier. I also would like to acknowledge the opportunity of being a member of the GGNB Times Newsletter editorial board for three years.

I would like to thank Maria for being the best flat mate ever. Thanks for the countless chats at the kitchen table or on the balcony but mostly for becoming my family here in Göttingen. And thank you for proofreading the thesis! I thank the rest of 'Der Harte Kern' - Lisa and Tina, and Jule for being by my side since my beginnings here in Germany. I'm very thankful for all the nice memories we share. A big thanks goes to Britta for the many hours of Volleyball that got me out of the lab and library. And to Christian - Thank you for all the wonderful distractions and for making the last months - despite writing the thesis - so special.

Agnes, Theresa, Laura and Alexandra, I thank you so much for your long-standing friendship, for making Austria a home to me and for your genuine interest in what life will bring next for me.

Meiner Familie gebührt der größte Dank. David - Ich bin unglaublich froh einen Bruder an meiner Seite zu haben und ich bin extrem stolz auf dich! Mama und Papa - eure Unterstützung und die Möglichkeit zu studieren hat diese Arbeit erst möglich gemacht. Danke für euer Vertrauen in meine Entscheidungen, dass ihr jede einzelne davon unterstützt und das Wissen, dass ich immer heim kommen kann. Diese Arbeit ist euch gewidmet.

Table of Contents

1. Summary ..... - 1 -
2. General Introduction ..... - 3 -
2.1. Development, function and evolution of body structures are governed by tightly regulated gene expression ..... - 3 -
2.2. Thesis overview ..... - 7 -
2.3. Comparative gene expression studies in development ..... - 8 -
2.3.1. *Schistocerca gregaria* as a model to study the role of pleuropodia in insect embryogenesis. - 10 -
2.4. Comparative gene expression studies in phenotypic evolution ..... - 13 -
2.4.1. *Drosophila melanogaster* as a model species to study head size and shape evolution ..... - 15 -
2.4.2. Mechanisms underlying context dependent gene expression divergence ..... - 19 -
3. Chapter I - Transcriptomics supports that pleuropodia of insect embryos function in degradation of the serosal cuticle to enable hatching ..... - 22 -
3.1. Abstract ..... - 25 -
3.2. Introduction ..... - 26 -
3.3. Results ..... - 28 -
3.3.1. Development of pleuropodia in the course of *Schistocerca* embryogenesis ..... - 28 -
3.3.2. Generation of a comparative RNA-seq dataset from developing pleuropodia and legs of *Schistocerca* ..... - 33 -
3.3.3. Identification of genes upregulated in the intensively secreting pleuropodia ..... - 35 -
3.3.4. The pleuropodia upregulate genes for cuticular chitin degrading enzymes ..... - 39 -
3.3.5. Pleuropodia upregulate transcripts for some proteases that could digest a cuticle ..... - 42 -
3.3.6. Pleuropodia are enriched in transcripts for immunity-related proteins ..... - 44 -
3.3.7. The pleuropodia do not upregulate the pathway for ecdysone biosynthesis ..... - 46 -
3.4. Discussion ..... - 47 -
3.4.1. Pleuropodia of *Schistocerca* express genes for the “hatching enzyme” ..... - 47 -
3.4.2. Pleuropodia in some other insects could secrete the “hatching enzyme” and their function may also vary among species ..... - 48 -
3.4.3. The pleuropodia of *Schistocerca* are enriched in transcripts for enzymes functioning in immunity - 49 -
3.4.4. Conclusions ..... - 49 -
3.5. Material and Methods ..... - 50 -
3.5.1. Insects ..... - 50 -
3.5.2. Description of embryonic stages ..... - 50 -
3.5.3. Immunohistochemistry on paraffin sections ..... - 50 -
3.5.4. Transmission (TEM) and scanning (SEM) electron microscopy ..... - 51 -

3.5.5. Preparation of the reference transcriptome ... - 51 -
3.5.6. Sequence analysis ... - 53 -
3.5.7. RNA-seq expression analysis ... - 53 -
3.5.8. GO enrichment ... - 55 -
3.5.9. Real-time RT-PCR ... - 55 -
3.6. List of abbreviations ... - 55 -
3.7. Data availability ... - 56 -
3.8. Competing interests ... - 56 -
3.9. Funding ... - 56 -
3.10. Author's contributions ... - 56 -
3.11. Acknowledgements ... - 56 -
3.12. Supplementary Figures ... - 57 -
3.13. Supplementary Tables ... - 65 -
4. Chapter II - Variation in a pleiotropic regulatory module drives evolution of head shape and eye size in *Drosophila* ... - 110 -
4.1. Abstract ... - 112 -
4.2. Introduction ... - 113 -
4.3. Results ... - 115 -
4.3.1. *Drosophila melanogaster* and *D. mauritiana* exhibit differences in dorsal head shape ... - 115 -
4.3.2. Difference in the transcriptomics landscape recapitulate observed morphological differences between *D. melanogaster* and *D. mauritiana* ... - 118 -
4.3.3. Central transcription factors regulate differentially expressed genes ... - 119 -
4.3.4. Pannier regulates genes that are differentially expressed between *D. melanogaster* and *D. mauritiana* ... - 121 -
4.3.5. Pnr activates and represses target genes in the eye-antennal disc ... - 124 -
4.3.6. Pannier and its co-repressor U-shaped participate in the same regulatory network during eye- and head development in *Drosophila* ... - 126 -
4.3.7. Overexpression of *pannier* phenocopies aspects of the differences observed between *D. melanogaster* and *D. mauritiana* ... - 129 -
4.4. Discussion ... - 132 -
4.4.1. A developmental model for natural variation in head shape and eye size ... - 132 -
4.4.2. Pnr and Ush represent a functionally linked pleiotropic module in the GRN underlying head and eye development ... - 134 -
4.4.3. GRN rewiring facilitates natural variation in pleiotropic developmental factors ... - 136 -
4.4.4. Evolution of GRNs and implications for convergent evolution of head shape and eye size ... - 137 -
4.4.5. Conclusion and Outlook ... - 139 -
4.5. Material and Methods ... - 139 -
4.5.1. Generation of the transcriptomic dataset ... - 139 -

4.5.2. Generation of the ATAC-seq dataset ..... - 141 -
4.5.3. Bioinformatics processes of the ATAC-seq dataset ..... - 142 -
4.5.4. Definition of a Pnr target gene list ..... - 142 -
4.5.5. qPCR ..... - 143 -
4.5.6. Antibody staining ..... - 144 -
4.5.7. Geometric Morphometrics ..... - 145 -
4.5.8. Overexpression/Knock-down of *pnr* and *ush* ..... - 145 -
4.5.9. *pnr* expression and lineage. ..... - 146 -
4.5.10. Immunostaining and imaging. ..... - 146 -
4.5.11. Adult head cuticle preparation. ..... - 146 -
4.5.12. Ommatidia Counting ..... - 146 -
4.6. Supplementary Figures ..... - 148 -
4.7. Supplementary Tables ..... - 162 -
5. Chapter III - Regulatory Divergence in the *Drosophila melanogaster* subgroup ..... - 169 -
5.1. Introduction ..... - 171 -
5.2. Results ..... - 173 -
5.2.1. Regulatory Divergence in the *D. melanogaster* subgroup ..... - 173 -
5.2.2. A comparative ATAC-seq dataset of three closely related *Drosophila* species ..... - 176 -
5.2.3. Genes with species specific regulatory regions are more often regulated in *cis* ..... - 178 -
5.2.4. Regulatory regions of genes, diverging in cis, show a higher sequence divergence ..... - 180 -
5.2.5. Regulatory divergence in transcription factors ..... - 182 -
5.3. Discussion ..... - 182 -
5.3.1. Regulatory divergence is context dependent ..... - 182 -
5.3.2. *cis*- regulatory divergence is due to changes in chromatin accessibility and sequence divergence - 185 -
5.3.3. Compensation and conservation of gene expression ..... - 187 -
5.3.4. Technical and other considerations ..... - 188 -
5.4. Conclusion ..... - 189 -
5.5. Material and Methods ..... - 190 -
5.5.1. RNA-seq ..... - 190 -
5.5.2. ATAC-seq ..... - 190 -
5.5.3. Conversion of Coordinates ..... - 192 -
5.5.4. Comparison of peak architectures ..... - 193 -
5.5.5. Sequence alignments ..... - 193 -
5.5.6. Overlap with DroID database ..... - 194 -
5.6. Supplementary Figures ..... - 195 -
5.7. Supplementary Tables ..... - 200 -

5.8. Appendix ..... - 203 -
6. General Discussion and Outlook ..... - 207 -
6.1. Integration of different datasets in comparative biological studies ..... - 207 -
6.2. Comparative gene expression studies and gene regulatory networks in development ..... - 210 -
6.3. Evolution of gene regulatory networks ..... - 214 -
7. References ..... - 220 -

List of Abbreviations

(qRT) PCR	(quantitative real time) polymerase chain reaction
3D	3-dimentional
AEL	after egg laying
agouti-related peptide2	*agrp2*
Ance	*Angiotensin-converting enzyme*
ASE	allele specific expression
ATAC-seq	Assay for Transposase-Accessible Chromatin
ato	*atonal*
BMP4	bone morphogenetic protein 4
bp	base pair
CaM	calmodulin
CDS	coding sequence
ChIP-seq	Chromatin Immuno Precipitation - sequencing
CHT	chitinase
CpG	C-phosphate-G
Ct	Cut
DEG	differentially expressed gene
DF	dorsal frons
DHS	DNase hyperactive sites
dl	*dorsal*
dll	*distal-less*
DNA	deoxyribonucleic acid
dpp/Dpp	*decapentaplegic*/Decapentaplegic
EC1/EC2	first and second embryonic cuticle
EcR	Ecdysone receptor
EGFR	Epidermal growth factor receptor
Evo-Devo	Evolutionary Developmental Biology
ey/Ey	*eyeless*/Eyeless
eyg	*eyegone*
F1 hybrid	filial 1 hybrid
FAIRE-seq	Formaldehyde-Assisted Isolation of Regulatory Elements
GAL	galactose
GO	gene ontology
GRN	gene regulatory network
GWAS	genome-wide association study
Hth	Homothorax
i.e.	id est
in. prep.	in preparation
JNK	c-Jun N-terminal kinases
Jra	Jun-related antigen
LEG	hind legs
lncRNA	long non-coding RNA
Mc1r	Melanocyte-stimulating hormone receptor

Mef2	Myocyte enhancer factor 2
MF	morphogenetic furrow (OR moulting fluid in **Chapter I**)
miR-92a	*micro RNA 92a*
miRNA	micro RNA
mRNA	messenger RNA
NAG	β-N-acetyl-hexosaminidase
Nej	Nejire
NGS	next generation sequencing
OC	orbital cuticle
oc	*ocelliless*
Pax6	Paired box protein 6
Pc	Polycob
PC	principal component
PCA	principal component analysis
Pitx1	*Paired Like Homeodomain 1*
PLP	pleuropodium
pMad	*phosphorylated Mothers against dpp*
pnr/Pnr	*pannier*/Pannier
QTL	quantitative trait locus
RNA	ribonucleic acid
RNAi	RNA interference
RPKM	Reads Per Kilobase Million
SC	serosal cuticle
sd	*scalloped*
SEM	scanning electron microscopy
Sfmbt	Sex comb on midleg-related gene containing four mbt domains
SNP	single nucleotide polymorphism
So	Sine Oculis
svb/Svb	*shavenbaby*/Shavenbaby
TEM	transmission electron microscopy
TF	transcription factor
Tin	Tinman
trn	*tartan*
tsh	*teashirt*
TSS	transcription start site
Ttk	Tramtrack
UAS	upstream activation sequence
wg/Wg	*wingless*/Wingless

1. Summary

The striking diversity in adult morphologies is the result of millions of years of adaptation of species to different environments and habitats. Fixed changes in populations or species are the consequence of mutations in the genome and thus in the developmental programs of body plans, their structures and organs. Years of studies in the field of 'Evo-Devo' have revealed that there exists only a limited number of genes, governing basic developmental processes, and that these so-called 'toolkit genes' are highly conserved even between distantly related species. It is nowadays accepted, that morphological diversification is often driven by changes in gene expression and subsequently the interplay of gene products. Since the expression of genes is tightly controlled in a spatiotemporal manner on several molecular levels, also the wiring of such gene regulatory networks is highly context dependent. Therefore, single cells, tissues and organs are characterized by a unique set of expressed transcripts and proteins which are specifically intertwined and govern their developmental programs. The advent of high throughput sequencing techniques provides nowadays the opportunity to analyze the transcriptome of developing structures in a highly specific manner and opens the possibility to understand how these toolkit genes are differentially used and rewired in different developmental and evolutionary contexts. In Chapter I of this thesis, I studied gene expression in a developmental context, using the emerging model species *Schistocerca gregaria* to understand the development and function of pleuropodia - small glandular structures forming on the first abdominal segment of many insect embryos. In Chapter II, I used a comparative transcriptomic dataset of developing eye-antennal discs in two closely related species of the *Drosophila melanogaster* subgroup to study the molecular basis of evolution of complex traits. The size and shape of the compound eyes and head structures vary extensively between *D. melanogaster* and *D. mauritiana* and show a typical trade-off between eye-size and head width. I could show that differential expression of *pannier* (*pnr)* underlies natural variation of eye size, ommatidia number and head width between these two species. In Chapter III, I combined an allele specific expression dataset of F1 hybrids between *D. melanogaster vs*. *D. mauritiana* and *D. simulans vs. D. mauritiana* with a newly generated comparative ATAC-seq dataset, to study gene expression divergence and sought to recapitulate the observed patterns in terms of nucleotide turnover and accessibility of regulatory regions. In summary, this works shows that

the combination of methods and various datasets allows to gain major insights into development, function, and evolution of morphological traits.

2. General Introduction

2.1. Development, function and evolution of body structures are governed by tightly regulated gene expression

The information how we and all other organisms develop, function and interact with our environment is encoded in our DNA which lies tightly packed as chromosomes in the nuclei of each of our cells (Figure 1A). During a process called transcription the genetic information encoded in genes is transcribed into messenger RNA (mRNA). The mRNA provides the template for the translational machinery, which translates the mRNA into amino acid sequences and eventually functional proteins (Figure 1C).

A typical eukaryotic gene locus is composed of several elements. The protein information is encoded in one or more exons, which together form the coding region (CDS), and are separated by introns. Transcription is initiated by the assembly of a basal transcription machinery at the promoter region, mostly located 5' upstream, close to the transcription start site (TSS) of the respective gene. This protein complex recruits the RNA polymerase that synthesizes the pre-mRNA. Where, when and how strong a gene is transcribed is though in the first place controlled by regulatory intronic or intergenic DNA regions, so called enhancers or *cis*-regulatory regions ((Davidson, 2001; Wray, 2003), Figure 1C). Therefore the respective genomic regions must be depleted of nucleosomes, which otherwise confer tight DNA packing. Hence, regulatory sequences must be accessible for transcription factors (TFs) that physically interact with the DNA by recognizing sequence-specific TF-binding motifs. This in turn leads to recruitment of additional TFs and co-factors. Enhancer sequences, are often of modular nature, meaning that several, locally separated regulatory regions modulate the expression of a single gene (e.g. (Adachi et al., 2003; Davidson, 2001; Stanojevic et al., 1991)). The advances in high throughput sequencing methods nowadays allow to reliably define the location of open chromatin regions in the genome. Approaches like ChIP-seq (Johnson et al., 2007; Robertson et al., 2007), FAIRE-seq (Giresi et al., 2007) or ATAC-seq (Buenrostro et al., 2013) are frequently used to define putative regulatory regions and allow to link them to gene expression, if combined with other methods like RNA-seq (Wang et al., 2009). However, how exactly enhancers carry out their regulatory function is not yet completely understood and different models of enhancer function have been proposed (Buffry et al., 2016). Chromosome

conformation capture methods combined with high throughput sequencing such as Hi-C (van Berkum et al., 2010) allow resolving the 3-dimensional chromatin states and are used to study how distantly located regulatory sequences exert their regulatory function (Furlong and Levine, 2018).

Each cell type of an organism is characterized by a certain combination of expressed genes and the defined interplay of their gene products. Since different cell types have to carry out distinct functions for a long period of time (depending on the life span of an organism), this function is ensured by tissue or even cell-specific gene expression (Lübbe and Schaffner, 1985). Traditional methods to quantify the expression levels of single genes include quantitative real-time PCR (qRT PCR,(Bustin, 2000)) and Northern Blotting (Alwine et al., 1977). The spatial distribution of transcripts can be studied by *in-situ* hybridization (Pardue and Gall, 1969). Nevertheless, only the advent of next generation sequencing (NGS) like RNA-seq facilitated the efficient genome wide assessment of gene expression by quantifying the complete mRNA content that is expressed at a certain time point in a cell or tissue (Wang et al., 2009). Disturbance of gene expression, and thus function, eventually leads to disease or death of the respective organism (e.g. (Dermitzakis, 2008; Emilsson et al., 2008)). For instance, in humans, the formation and progression of cancer is tightly linked to aberrant gene expression and regulation (e.g. (Liang and Pardee, 2003)). Therefore, the expression of genes has to be under tight spatial and temporal regulation, which is ensured on several molecular and cellular levels (Figure 1C). The accessibility of regulatory regions for instance is highly dependent on the tissue and developmental stage (e.g. Bozek et al., 2019). Furthermore, biochemical modifications of DNA (methylation) and histone proteins (methylation, acetylation, phosphorylation and many others) influence gene expression (Kouzarides, 2007; Lawrence et al., 2016) (Figure 1C). In *Drosophila* dosage compensation relies for example on the acetylation of lysine 16 residues on the H4 histones of the X-chromosome, allowing the increase of transcription in males by decondensation of the chromosomes (e.g. (Akhtar and Becker, 2000; Turner et al., 1992)). Additionally, methylation of Cytosines has been linked to repression of transcription (reviewed in Bird and Wolffe, 1999). In vertebrates for example, promoter or enhancer regions, often containing so-called CpG-islands are usually depleted of methylated CpGs and hyperacetylated histones, marking actively transcribed genes.

The spatially and temporally restricted availability of TFs and co-factors that bind to accessible regulatory regions further represents a level of context specific gene regulation. One example of transcriptional co-regulation, which will be introduced in **Chapter II** in more detail can be found in the developing *Drosophila* wing disc. Pannier (Pnr), a GATA transcription factor which usually activates expression of its target genes, interacts in a spatially defined manner with U-shaped (Ush) (Fromental-Ramain et al., 2010, 2008). The resulting heterodimer loses the activating role of Pnr but acquires a repressing function (Haenlin et al., 1997). Also, post-transcriptional processes can modulate gene expression in a context dependent manner. For instance, the context dependent expression of small regulatory RNA molecules such as microRNAs (miRNAs) modifies the stability of mRNA or the efficiency with which an mRNA molecule is translated (reviewed in Bartel, 2018; Kittelmann and McGregor, 2019). Also, for long-non-coding RNAs (lncRNAs) it has been established that they are transcribed in a highly spatially and temporally controlled manner and are suggested to influence for example the expression of genes in their close genomic vicinity (Kopp and Mendell, 2018; Ponting et al., 2009; Sarropoulos et al., 2019). These are only few of the many examples that show that tissue and stage specific gene expression is orchestrated on different levels of the gene regulation machinery.

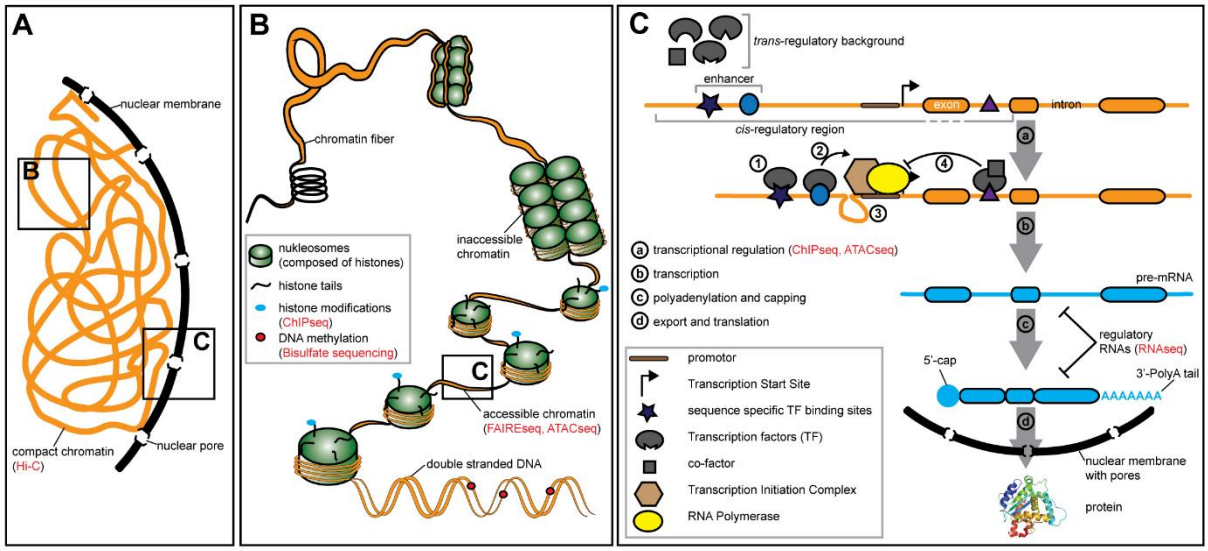


Figure 1. Gene expression is tightly controlled. A. The DNA lies heavily packed as so-called chromatin in the nuclei of eukaryotic cells. **B.** Formation of chromatin is carried out by wrapping DNA around histones, which are composed of nucleosomes. Regions of loose packing, characterized by nucleosome depletion, are in general more accessible for transcription factors (TFs) and loci in these regions are mostly actively transcribed. In contrast, tightly packed DNA is inaccessible to regulatory proteins and subsequent transcription. Biochemical modification of histones or cytosines provide another level of gene regulation. **C**. A eukaryotic gene locus is composed of one or more exons, which make up the CDS of the gene. Regulatory regions are located in introns, separating the exons, or in intergenic regions. Transcription is initiated at the promoter region, 5' upstream of

the transcription start site (TSS), and TFs bound to enhancer regions further regulate gene expression. The figure is taken from (Buchberger et al., 2019).

While gene expression has to be tightly controlled to ensure proper organ development and function, many evolutionary studies revealed that divergence in gene expression is a key driver for phenotypic evolution (Alvarez et al., 2015; Carroll, 2005; King and Wilson, 1975; Todd et al., 2016). One of the most classical examples, where differences in morphologies were associated with differential gene expression is the work of Abzhanov and colleagues, who linked higher expression of bone morphogenetic protein 4 (BMP4) to wide beak morphology in ground finches (Abzhanov, 2004), whereas development of long beaks of cactus finches is mainly driven by higher levels of calmodulin (CaM) (Abzhanov et al., 2006). In East African cichlid fish it has recently been revealed, that changes in the expression of the *agrp2* gene, defines the pigmentation pattern of different radiations (Kratochwil et al., 2018). Similarly, adaptive changes in abdominal pigmentation of African *Drosophila* populations are caused by expression variation of the *ebony* gene (Pool and Aquadro, 2007; Rebeiz et al., 2009). Changes in gene expression levels could be due to changes in a gene's own regulatory regions (*cis*-regulatory divergence) or due to divergence of upstream regulators, such as transcription factors or regulatory RNAs (*trans*-regulatory divergence) (Cowles et al., 2002; Wittkopp et al., 2004). For many simple traits, including pigmentation, trichome formation or loss of specific skeletal structures, it has been shown that the causative underlying mutations are often located in the non-coding, regulatory regions of the locus (e.g. Chan et al., 2010; McGregor et al., 2007; Prud'homme et al., 2006; Rebeiz et al., 2009), which would eventually affect the expression of the respective gene. If this also applies to quantitative, complex traits like size and shape of organs and structures remains to be established.

In summary, gene expression is a central biological process that transfers the information stored in the genome of an organism to its development, function and evolution.

2.2. Thesis overview

During my PhD work I applied comparative gene expression studies to gain new insights into:

I. developmental processes and organ function,
II. the evolution of complex morphological traits and
III. molecular mechanisms underlying gene expression divergence.

Chapter I **'Transcriptomics supports that pleuropodia of insect embryos function in degradation of the serosal cuticle to enable hatching'** resulted from a collaboration with Dr. Barbora Konopová and Dr. Alastair Crisp. Applying comparative RNA-seq, we provide strong evidence that pleuropodia in the locust *Schistocerca gregaria* indeed participate directly in the digestion of the serosal cuticle during embryogenesis and reveal that they also might take over a role in insect immunity.

In Chapter II **'Variation in a pleiotropic regulatory module drives evolution of head shape and eye size in *Drosophila*'** I studied differences in gene expression dynamics between *D. melanogaster* and *D. mauritiana* and show that differential expression of the conserved transcription factor Pnr underlies variation in head shape and ommatidia number between the two species. Additionally, I found that the co-factor of Pnr, Ush is expressed and functional in the developing eye-antennal discs of *Drosophila* and therefore represents a new player in the eye and head GRN.

For Chapter III **'Regulatory divergence in the *Drosophila melanogaster* subgroup'** I combined previous knowledge about regulatory divergence in three species of the *D. melanogaster* subgroup with a newly generated ATAC-seq dataset to study if patterns of *cis*- and *trans*-regulatory divergence can be recapitulated on the basis of open and accessible chromatin regions.

In the following paragraphs I will provide an overview of the current knowledge to introduce each of the three chapters.

2.3. Comparative gene expression studies in development

The goal of molecular studies in developmental biology is to understand how gene products work together to provide instructive signals that control developmental processes (Wolpert and Tickle, 2011). The context dependency of gene expression ensures that specific cell types and tissues are characterized by the expression of a unique set of transcripts which are then translated into transcription factors and structural proteins, making up the building blocks of the respective cell, tissue and organ. Assessing and comparing mRNA composition and gene expression levels across developmental time points provides therefore the chance to better understand the molecular underpinnings of developmental processes and eventually organ functions.

Much of our detailed knowledge about the genes coordinating developmental processes in insects is deduced from studies in the model species *D. melanogaster*. In this species, for instance, the establishment of the body axis, was first studied and understood in great detail: The translation of maternally deposited mRNA leads to the activation of a hierarchical gene activation cascade and subsequently to anterior-posterior segmentation of the complete developing embryo (e.g. Johnston and Nüsslein-Volhard, 1992). Since the advent of RNA-seq, major effort has been made to characterize not only the location and role of single genes, but to establish a complete catalog of transcripts and their expression dynamics in developing and adult tissues (e.g. Brown et al., 2014; Graveley et al., 2011). One major drawback of focusing developmental studies on established model systems is that derived structures or organs that are not present in the vinegar fly are less well studied and understood. Easy accessibility and constant reduction of costs for next generation sequencing techniques nowadays allow to explore the development and function of single organs in nearly every species, including plants and animals (Wang et al., 2009) and has greatly expanded the use of emerging model organisms in developmental biology (Ellegren, 2014).

Since genomic or transcriptomic resources are usually sparse in emerging model systems, the first step often includes the *de-novo* assembly of reference genomes or transcriptomes against which the short reads can subsequently be mapped (reviewed in Cheng et al., 2018). Depending on the species, *de-novo* transcriptome assembly can be achieved with the help of a reference genome, or if not available by using the short reads directly for assembly (Cheng et al., 2018). Blasting the *de-novo* assembled transcriptome against databases like

UniProt/Swiss-Prot, allows to assign putative functions to transcripts and by this allows to retrieve such information also for transcriptome datasets of non-model species ("UniProt," 2019).

Once references are established, genome wide gene expression can be compared across different conditions of interest, which can include the comparison of different stages during development of a certain organ, or the comparison of different tissues. Such an analysis usually results in long lists of differentially expressed genes. Depending on the exact research question, a major challenge is to extract meaningful information from such large datasets. A first helpful step is often the integration of prior molecular, cellular or functional knowledge. This information can be retrieved from the Gene Ontology (GO) database, which links a particular gene to its function by annotating it to one or more defined GO-terms. Using a statistical framework, it allows to understand in which molecular functions, biological processes and cellular components differentially expressed genes are enriched in (Ashburner et al., 2000; The Gene Ontology Consortium, 2019). If, for instance, different developmental stages are studied, the expression dynamics can be analyzed in more detail by clustering genes that share a similar expression profile. It has been proposed, that such co-expressed genes are often co-regulated by the same upstream transcription factors and are involved in related biological functions (Yu et al., 2003). Following that assumption, clustering algorithms that group genes with similar expression levels over a certain period of time, combined with GO-enrichment analysis and an upstream search for transcription factor binding motifs, provides a meaningful tool for the reconstruction of developmental gene regulatory networks (GRNs). Note that the direct search for transcription factor binding motifs works well for established model systems, where databases of TF binding profiles exist. These include e.g. humans and mouse as representatives of vertebrates, *D. melanogaster* for insects or *Arabidopsis thaliana* representing plants (Khan et al., 2018). Nevertheless, for non-model systems a so-called *de-novo* motif search can be useful to find overrepresented motifs in regulatory regions of co-expressed genes, followed by a subsequent comparison to known motifs (e.g. Bailey et al., 2009), since transcription factor binding domains are often conserved along large phylogenetic distances.

Studying the development and function of organs in classical model organisms like *Drosophila* has brought major insights in many aspects of biology. However, for some questions

in developmental or evolutionary biology, the selection of a handful of established model species does not necessarily represent the best systems. Developmental processes that are highly derived in *Drosophila* are for example insect head development (Davis and Patel, 2002; Grossniklaus et al., 1994) or the embryonic development of insects. Extraembryonic membranes that usually protect insect eggs from desiccation have been secondarily reduced in higher flies (Schizophora) (Glaser-Schmitt and Parsch, 2018; Jacobs et al., 2013; Schmidt-Ott, 2000) and certain structures which play a role during hatching of the embryo, like pleuropodia (see next section) are missing in the model species *Drosophila*. Studying traits which are not present in classical model species requires therefore to establish morphological and genomic resources in a variety of species. We applied a comparative RNA-seq approach to pleuropodia and leg buds of the desert locust *Schistocerca gregaria* (*S. gregaria*) and combined this with a thorough description of their ultrastructure throughout development to understand their function during insect embryogenesis. We further provide a transcriptomic resource to understand appendage differentiation by comparing two serially homologous structures.

2.3.1. *Schistocerca gregaria* as a model to study the role of pleuropodia in insect embryogenesis.

Insects are the most species-rich animal group on this planet and their success is the result of several evolutionary specializations which allowed them to conquer all environments such as air, water and land (Losos, 2014). These include for instance the emergence of wings in pterygotes (winged insects), the development of three life stages in holometabolous species or eusociality in several insect lineages (Losos, 2014). The colonization of land also required protection against desiccation, especially during embryonic development (Jacobs et al., 2013). Most insect embryos possess two membranes, the amnion and the serosa (Figure 2B), which do not directly contribute to the formation of the insect body, but often cover the entire embryo and take part in crucial developmental processes including - amongst many others - cuticle production, immune responses, or hatching (e.g.(Jacobs et al., 2015, 2014, 2013; Panfilio, 2008)). A non-cellular, three-layered serosal cuticle, which is secreted by the serosa itself, lies between this non-embryonic membrane and the eggshell ((Goltsev et al., 2009; Jacobs et al., 2015) Figure 2B). The serosal cuticle has to be digested prior to hatching of the embryo. In grasshoppers and glowworms for instance, the two inner layers of the serosal cuticle merge and decay shortly before hatching, whereas the eggshell and the remaining layer of the

serosal cuticle layer is mechanically torn by the mandibles (H. Slifer, 1937; Kobayashi et al., 2003). The pair of pleuropodia develops in a plethora of insects at the first abdominal segment of the embryo (Figure 2A; (Wheeler, 1890)) and degenerates at the end of embryogenesis. Orthopterans have proven to be a valuable model to study these small organs, since they are - due to their large embryos - easily accessible. Consequently, it was already shown 80 years ago in grasshoppers, that pleuropodia are involved in the digestion of the serosal cuticle (H. Slifer, 1937; Slifer, 1938). However, the clear mechanism how these organs are involved in this process has remained elusive so far. It was suggested, that they facilitate digestion indirectly via secretion of the ecdysone hormone (Novak and Zambre, 1974), or directly by secreting the so-called 'hatching enzyme' (H. Slifer, 1937; Louvet, 1975). Up to now, a thorough description of pleuropodia development, their function and transcriptomic landscape is still missing. Furthermore, since pleuropodia are serially homologous to leg buds, these two structures provide an excellent model to study when and how initially similar structures differentiate during the development of an organism. The proper development of body structures is highly dependent on tissue and stage specific gene expression and the correct interplay of the translated proteins. The methods described in the latter section nowadays allow to generate relatively easily transcriptomes from different organs of non-model species, taking the spatiotemporal gene expression into account. With this they provide the basis for comparative gene expression approaches, which permit to recapitulate and understand the developmental programs of differentiating, serially homologous organs.

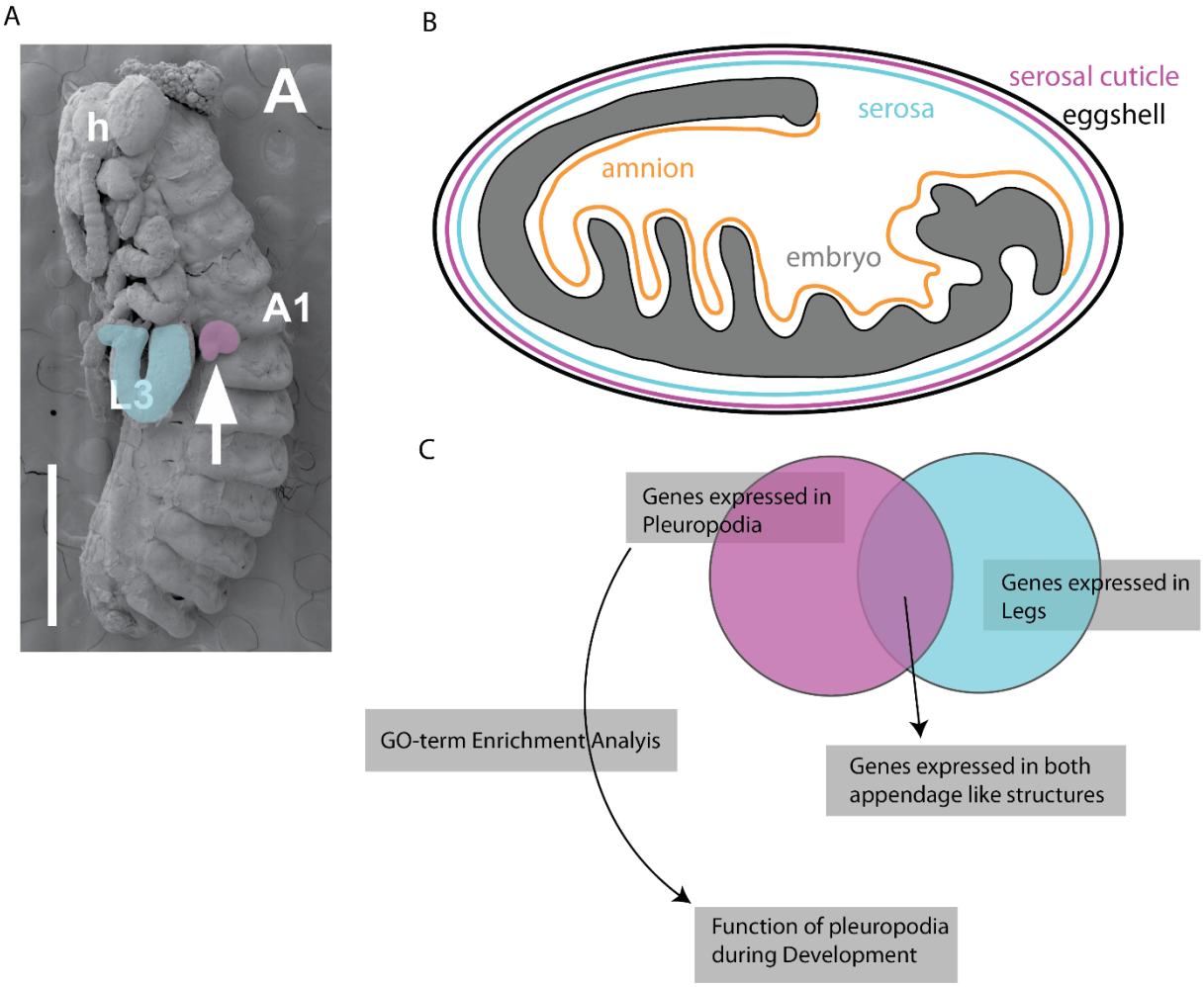


Figure 2. Pleuropodia and their role during insect embryogenesis. A. A pair of pleuropodia develops at the first abdominal segment in insect embryos (here marked with the white arrow and pink labelled in an embryo of *S. gregaria*). Pleuropodia and the third leg pair (in blue) were dissected to generate a comparative transcriptomic dataset (adapted from (Konopová et al., 2019)). **B.** Schematic representation of an insect embryo (germband stage). The embryo (in grey) is covered by the amnion (in orange). The serosa (in blue) envelopes the complete embryo and secretes the serosal cuticle (in pink) which lies between the serosa and the egg shell (in black) (after (Panfilio, 2008)) **C.** Experimental set-up of the comparative gene expression study to analyze function and putative new roles of pleuropodia during insect embryogenesis.

We therefore generated a comparative embryonic RNA-seq dataset of *Schistocerca gregaria* (*S. gregaria*) pleuropodia and legs (Figure 2 A and C), to investigate on a transcriptional level how the pleuropodia facilitate hatching of the embryo. The possibility to dissect pleuropodia and legs provided the opportunity to generate tissue specific datasets at 10 timepoints, also accounting for the temporal context dependency of gene expression. Combined with an in-depth morphological characterization, our results provide interesting insights into the development of pleuropodia, their function during hatching and putative roles in the embryo's immunity and are described in **Chapter I** of this thesis.

2.4. Comparative gene expression studies in phenotypic evolution

Besides far-reaching novelties, the adaptation to different environments is also facilitated by the ability to change the size and shape of organs and other body parts. The most classic example for natural variation in size and shape are the various beak forms of Galápagos finches, where changes in beak morphology were fundamental for the adaptation to different environments and food sources (Abzhanov, 2004; Abzhanov et al., 2006; Schluter, 2000). Morphological differences that are fixed across populations or species are the result of heritable changes in the genome (Figure 3). Even though this fact is widely accepted, pinpointing the exact molecular changes has been shown to be rather difficult and only few studies succeeded in resolving the causative genomic changes that underlie variation in adult phenotypes. This is mainly due to two reasons. First, variation in many traits, but especially complex traits like size and shape, are influenced by several genomic loci, i.e. they are polygenic (Boyle et al., 2017). Second, causative changes are not always found in the coding region of a gene (CDS), potentially changing the function of the resulting protein, but it is nowadays believed, that many changes occur in so-called *cis*-regulatory regions, affecting the expression of the respective gene (Wray, 2007).

While selection mostly acts on adult structures, developmental processes define the size and shape of the respective organ. Therefore, fixed changes in adult structures are the result of variation in developmental processes (Figure 3). By comparing the development of organisms one can thus reveal mechanisms underlying morphological divergence. The task of finding the genetic causes for phenotypic variation is usually addressed in the field of evolutionary developmental biology ('Evo-Devo'), the combination of evolutionary studies and developmental biology. 'Evo-Devo' aims to assess conserved aspects as well as differences in developmental programs between species that eventually result in variation in adult morphology (e.g.(Hall, 2003; Raff, 2000), Figure 3).

Figure 3. Genetic changes in the genome, which can occur in coding regions but also regulatory regions (light blue box) underlie changes in development by rewiring developmental gene regulatory networks (dark blue box) and subsequent variation in adult morphology (yellow box). If a certain phenotype provides an advantage in fitness in a specific environment (green box), these specific phenotypes will eventually be more common than others ('natural selection', grey box).

Numerous studies in this field resulted in exciting findings, such as the observation that a set of highly conserved transcription factors and signaling pathways governs the development of organisms over large phylogenetic distances from invertebrates to vertebrates. This was impressively shown in the case of HOX genes, a cluster of homeobox transcription factors, which define the anterior-posterior axes of all metazoans (Duboule and Dollé, 1989; Graham et al., 1989; McGinnis and Krumlauf, 1992; Scott and Weiner, 1984). Another well-described example is the *Pax6* gene, which is conserved in all organisms with light sensitive cells. Loss of function of this gene results in a no-eye phenotype in mouse embryos as well as in the vinegar fly *Drosophila*, where the gene was typically called *eyeless* (*ey*) (Hill et al., 1991; Quiring et al., 1994). The coding sequences of the two homologous proteins are strikingly similar, illustrated by the observation that the mouse protein can rescue mutants in the fly (Halder et al., 1995). Therefore, despite the diversity present in nature, the development of organisms is controlled by a limited set of highly conserved regulators, the so called 'developmental toolkit' (Carroll et al., 2001). Consequently, one major question in evolutionary biology is to understand how phenotypic diversity can arise in the light of generally highly conserved developmental regulators. In some cases, the causative mutations underlying phenotypic variation have been identified in protein coding sequences. Hoekstra and colleagues linked a fixed mutation in the gene, encoding for the receptor Mc1r, to differences in color patterns between subspecies of the beach mouse, *Peromyscus polionotus* (Hoekstra et al., 2006). Additionally, variation in HOX proteins has been shown to drive body plan diversification (Grenier and Carroll, 2000). However, many genetic variants identified for instance by quantitative genetics approaches mapped to non-coding regions (Dixon et al., 2007; Gilad et al., 2008; Jia and Xu, 2007). Already King and Wilson concluded in 1975 that much of the variation that can be observed between species, must be rather based on the way how genes are expressed than on changes in protein sequences themselves (King and Wilson, 1975). Therefore, variation in gene expression underlies phenotypic evolution. Here, we address the question how gene expression diverges in closely related species, and we use *Drosophila* head and eye development as a model to understand how body structures change their size and shape during evolution.

2.4.1. *Drosophila melanogaster* as a model species to study head size and shape evolution

Many studies assess the consequences of gene expression divergence by studying classical, discrete traits, like trichome patterns (e.g. McGregor et al., 2007), coloring patterns (e.g.(Gautier et al., 2018; Kratochwil et al., 2018; Prud'homme et al., 2006)) or the loss or gain of skeletal structures (Chan et al., 2010; Xie et al., 2019). However, in recent years, researchers also started to focus on the genomic basis underlying complex trait evolution, such as changes in size and shape of adult structures. The vinegar fly *D. melanogaster* but also its closely related sister species, *D. simulans* and *D. mauritiana* regularly serve as model species to study evolution of organ size. Hagen et al. showed for example that differences in the expression of *tartan* (*trn*) underlies the evolution of male genitalia size between *D. simulans* and *D. mauritiana* (Hagen et al., 2018). Especially variation in head and eye structures of *Drosophila* has been of particular interest in recent years (Arif et al., 2013a; Gaspar et al., 2019; Hilbrant et al., 2014; Keesey et al., 2019; Norry et al., 2000; Posnien et al., 2012). In comparison to its sister species, *D. melanogaster* has very small eyes with a broad interstitial face cuticle. In contrast, *D. mauritiana* has bigger eyes with a reduced face cuticle (Figure 4A and B, (Posnien et al., 2012)). Interestingly, it has recently been shown in a large-scale screen covering more than 60 *Drosophila* species that the trade-off between eye and head size is a common feature of *Drosophila* and most likely represents a functionally relevant subdivision of the visual and olfactory system (Keesey et al., 2019).

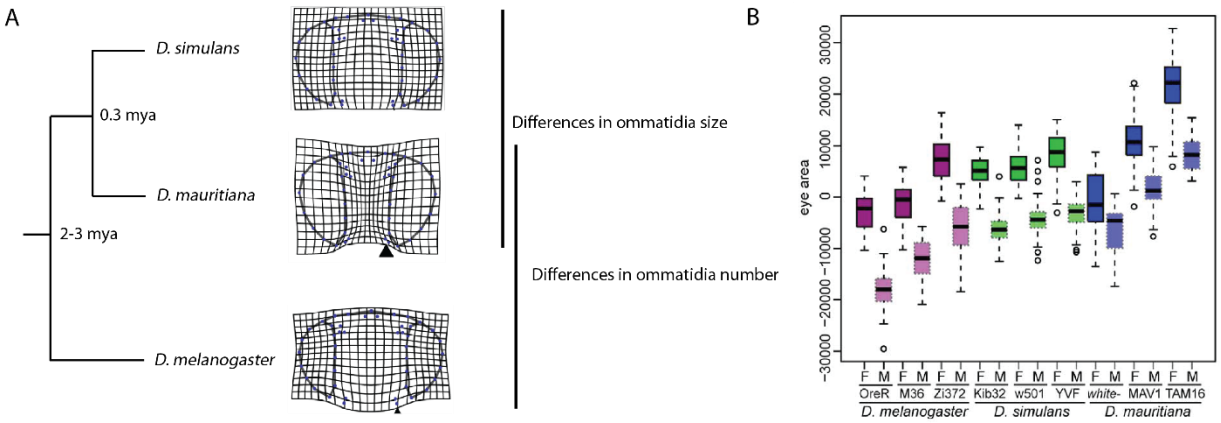


Figure 4. Natural Variation in eye size and head shape between closely related *Drosophila* species. A. Species in the D. *melanogaster* subgroup show extensive variation in eye size and head shape. They display the typical trade-off between the head capsule and the compound eye, where a larger eye area goes hand in hand with a narrower interstitial face cuticle and vice-versa. *D. melanogaster* has very small eyes, and a broad face, whereas *D. mauritiana* has very large eyes, which is especially pronounced in the dorsal part of the compound eyes.

Differences in eye size can either arise due to variation in ommatidia number, which is the case between *D. melanogaster* and *D. mauritiana*, or due to changes in ommatidia size, as observed for *D. mauritiana* vs. *D. simulans*. **B.** Eye area differences in different strains of *D. melanogaster*, *D. simulans* and *D. mauritiana*. OregonR (*D. melanogaster*) and TAM16 (*D. mauritiana*) show the most extreme phenotypes on both ends of the spectrum. Figure adapted from (Posnien et al., 2012).

The natural variation in *Drosophila* eye size and head shape provides an excellent model to study evolution of complex traits, since the GRNs that govern the development of these structures were already extensively studied. The *Drosophila* head develops from so-called eye-antennal imaginal discs which reside in the larva, attached to the mouth hooks and the two brain lobes. These paired epithelial cell sheets eventually give rise to several distinct adult head structures, including the head capsule, eyes, antennae and mouthparts (Haynie and Bryant, 1986). They have been used to study basic questions in developmental biology, including pattern formation, organ growth or the establishment of compartment boundaries (reviewed in Kumar, 2018). The developing eye-antennal imaginal disc grows homogeneously during the first two larval stages. Only at the end of the second instar the so-called morphogenetic furrow (MF) starts sweeping across the tissue, commencing at the posterior end of the disc. Cells in front of the MF stop dividing after a final mitotic wave. Cells posterior to the MF undergo a second and final round of mitosis, generating the cells, that make up each ommatidium, including for instance photoreceptors and cone cells (Wolpert and Tickle, 2011). Therefore, at the end of larval development the number of ommatidia in the adult compound eye is already defined.

All imaginal discs are formed by two layers, the disc proper and the peripodial epithelium. Both layers are connected via the cuboidal marginal cells ((Lim and Choi, 2004), reviewed in (Gibson and Schubiger, 2001; Kumar, 2018)). The squamous peripodial epithelium is defined by its large cell nuclei that can easily be distinguished from the columnar epithelial cells in the disc proper ((Auerbach, 1936), Figure 5A and B). The disc proper gives rise to most of the adult head structures, whereas the peripodial epithelium is thought to contribute to parts of the body wall cuticle (Figure 5C, (Fristrom et al., 1993; Milner et al., 1984)). It is nowadays accepted that the peripodial epithelium is essential for proper eye development, playing a role for instance in coordinating signaling pathways involved in dorsal-ventral patterning or MF progression, as well as disc growth via microtubule-based extension signaling through the lumen between the two layers (Gibson and Schubiger, 2000). This second epithelium is also important during pupal stages, where the two eye discs evert and finally fuse

to form the adult head structures. Mechanistic analyses suggested that the reduction of the peripodial epithelium area pushes the eye over the antennal area and by this facilitates morphogenesis of the head (Milner et al., 1984).

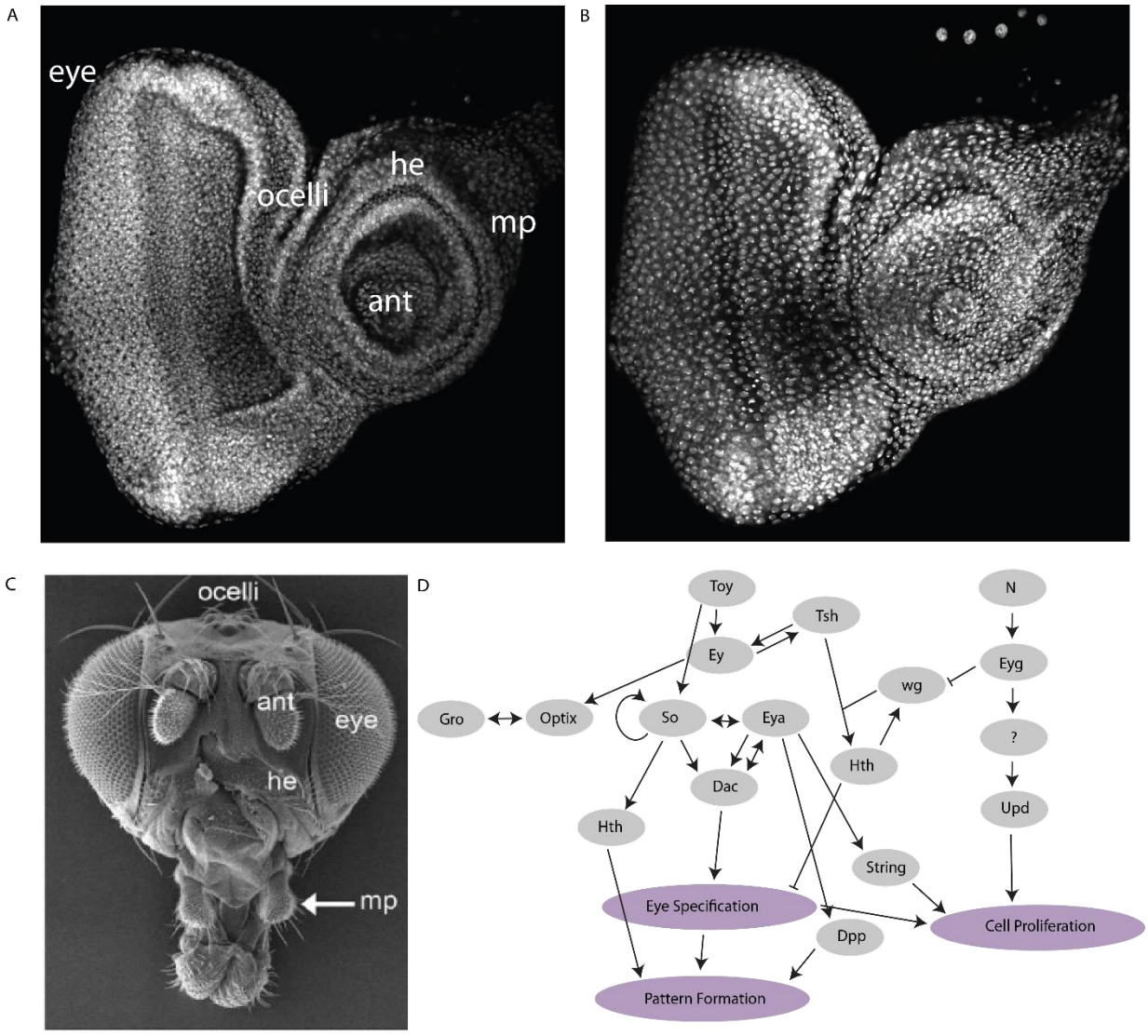


Figure 5. Eye and head development in *Drosophila*. A. In the third instar eye-antennal disc it can already be determined which part will give rise to which adult structure (eye, ocelli, head (he), antenna and maxillary palp (mp). **B.** Same eye-antennal disc as in A., focusing on the peripodial epithelium, characterized by large nuclei, stained with DAPI. **C.** Adult *Drosophila* head, the structures are labelled as in A. **D.** A simplified scheme of the GRN network governing eye development in *Drosophila* (Figure adapted from (Kumar, 2009)).

The GRN governing eye and head development is among the best studied in *Drosophila*. It is composed of a set of genes, the so-called retinal determination genes. On top of this cascade stands the famous *Pax6* homolog *ey* as a master regulator for eye development (Callaerts et al., 1997). Besides the retinal determination genes, important signaling pathways, including Wnt-, Dpp- and Notch signaling are part of the GRN and are involved in eye/head specification and cell proliferation (reviewed in (Kumar, 2009), Figure 5D).

The retinal determination genes get restricted to the posterior part of the developing eye-antennal disc during the second larval instar and by this stage *ey* is not expressed in the antennal part anymore. Instead, expression of the transcription factor Cut (Ct) can be detected in the anterior part of the disc. One important hallmark of this interplay of transcription factors is that they are not activated in a hierarchical cascade but interact in different GRNs which are themselves interconnected (Kumar, 2009; Treisman, 2013). These GRNs do not only include activation between transcription factors and their target genes but also involve feedback loops and repression of locally restricted GRNs: Wang and Sun showed that the expression of *ey* in the antennal part is repressed by Ct and Homothorax (Hth), whereas Sine oculis (So) is activated by Ey in the eye disc and represses Hth and So (Wang and Sun, 2012). Also, the growth of the final adult structures and therefore the size relationship among them are controlled via the repressing function of specific transcription factors. For instance, Wingless (Wg) signaling is important for defining the head cuticle fate by repressing retinal development and in turn promotes dorsal head specification (Magri et al., 2018; Treisman and Rubin, 1995). Therefore, in order to ensure the development of several functional organs and structures from one single epithelium, the underlying, intertwined GRNs must be tightly controlled and regulated.

Given the observed variation in eye size and head width within the *D. melanogaster* subgroup, we sought to study the evolution of this trade-off in *D. melanogaster* and *D. mauritiana* and focused on recapitulating where GRNs in closely related species evolve. Following the assumption that variation in gene expression is a major driver of phenotypic evolution, we generated a comparative transcriptomic dataset covering three distinct stages during eye and head development (72h after egg laying (AEL), 96h AEL and 120h AEL) in both species. Differential expression analysis together with a transcription factor binding site analysis showed that the GATA factor Pannier (Pnr) regulates many genes that are differentially expressed between *D. melanogaster* and *D. mauritiana*. We found that the transcript of *pnr* itself is differentially expressed in the two species during eye development. Additionally, our genome wide approach allowed us to characterize U-shaped (Ush), a co-factor of Pnr, as a previously unknown player in the GRN of the developing eye-antennal disc and could show that they genetically interact. Overall, we show in Chapter II that higher expression of *pnr* in *D. mauritiana* underlies part of the observed natural variation in eye size and head shape between *D. mauritiana* and *D. melanogaster*.

2.4.2. Mechanisms underlying context dependent gene expression divergence

While gene expression represents a great intermediate phenotype to study development and the molecular basis of phenotypic variation, it is also of major interest to gain comprehensive insights into the mechanisms underlying gene expression divergence itself. Divergent gene expression can arise due to two different mechanisms; either due to differences in the regulatory region of the differentially expressed gene itself (*cis*-regulatory divergence, Figure 6) or due to changes in an upstream regulator (*trans*-regulatory divergence, Figure 6) (Cowles et al., 2002; Wittkopp, 2005; Wittkopp et al., 2004). *cis*-regulatory divergence is the result of variation in a gene's regulatory region, caused by nucleotide changes in promoter or enhancer sequences that lead for instance to divergence in transcription factor binding (Wittkopp, 2013). *trans*-regulatory divergence is caused by changes in the upstream gene regulatory cascade, for instance in an upstream transcription factors, which would onsequently affect the transcriptional response following its binding to regulatory regions. Differences in the functionality of such an upstream factor can either be due to changes in the coding region, affecting for instance DNA-binding affinity, or due to differences in its expression, influencing the amount of available transcription factor in a given cell or tissue (Wittkopp, 2005). Even though *trans*-regulatory changes are mostly referred to as changes in transcription factors, it is noteworthy to mention, that upstream changes can occur on all levels of the upstream gene regulatory cascade, including for instance regulatory miRNAs (Figure 1).

Allele specific expression analysis (ASE) has been used to gain mechanistic insights into gene expression divergence. This approach makes use of an F1 hybrid generation by comparing gene expression in homozygous parent species with expression of their alleles in the same *trans*-regulatory background of the heterozygous hybrid ((Cowles et al., 2002; Wittkopp et al., 2008, 2004) Figure 6). Is a specific allele still differentially expressed in the hybrid background, then the causative mutation underlying differential expression of the respective genes in the parentals is thought to be located in the gene's own *cis*-regulatory region (Figure 6). If the two alleles do not show differential expression in the hybrids anymore, then the differential expression in the parental species is due to changes in upstream *trans*-regulatory factors, which are neutralized in the common hybrid background. The approach also reveals genes, whose expression is kept conserved in all conditions, i.e. neither the genes in the parental species, nor the alleles in the hybrid are differentially expressed. ASE also gives insights into compensatory

mechanisms, for instance when the expression of a gene is conserved in the parental species but the two allelic variants in the hybrid do show significant differential expression ((McManus et al., 2010) see Figure 6). Even though ASE is a valuable tool to gain genome wide insights into the mechanisms that underly gene expression divergence, the causative locus underlying differential gene expression cannot be directly inferred (reviewed in (Buchberger et al., 2019)). Additionally, one can only reveal changes over short evolutionary distances, since they rely on the ability of two parental species to produce viable hybrids.

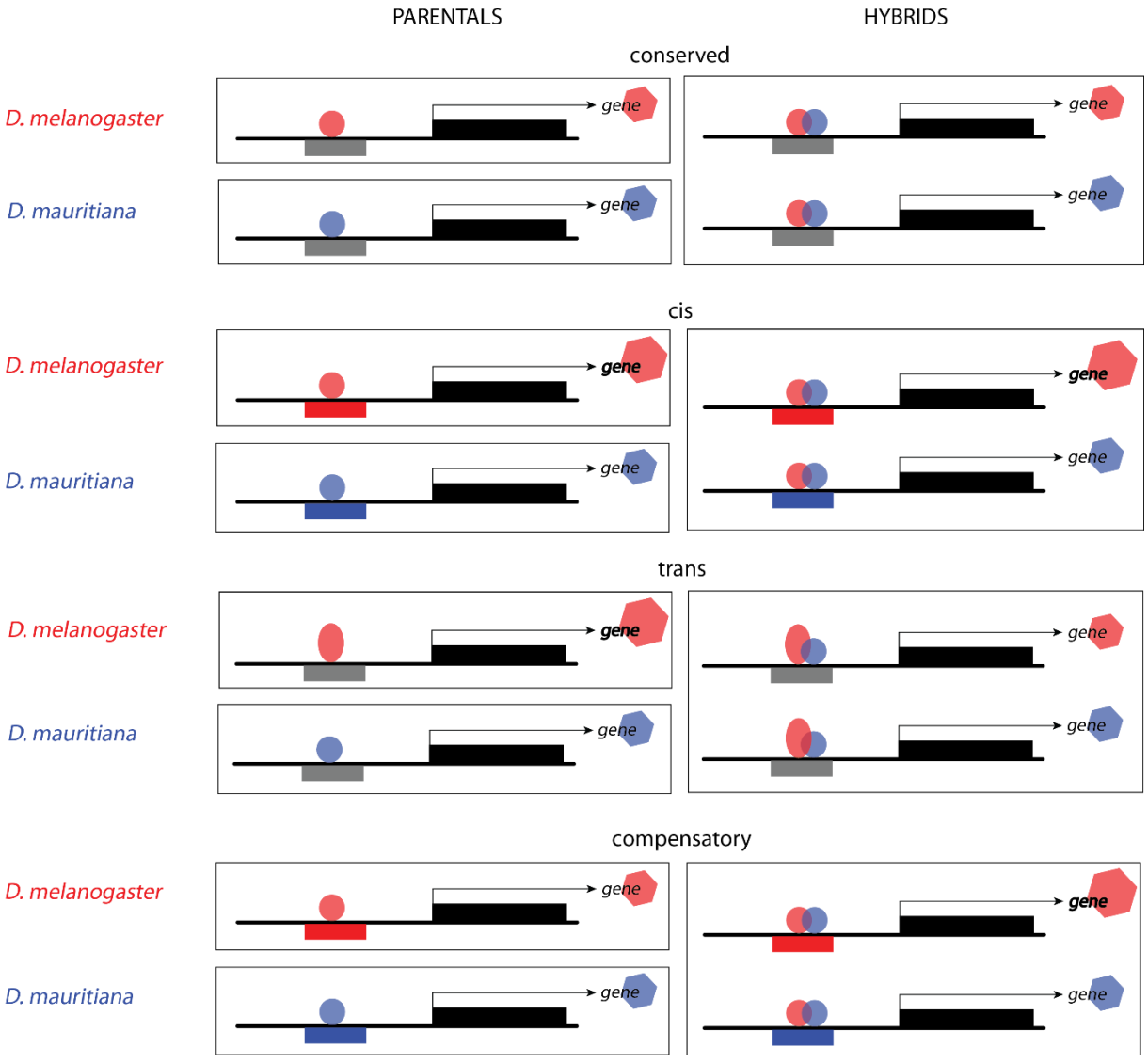


Figure 6. Allele specific expression analysis to study gene expression divergence. Parental species are shown on the left side: Red – *D. melanogaster* and blue – *D. mauritiana*. The colored bars represent the *cis*-regulatory elements of the respective alleles. In the F1 hybrid the *trans* background (TFs and co-factors) contains factors from both parents, therefore only differences in the *cis*-regulatory regions of the two alleles will influence differences in allelic expression. A gene is called 'conserved' if neither the genes in the parental species, nor the two alleles in the hybrids are differentially expressed. A gene is differentially expressed due to *cis*-regulatory changes, if it is higher expressed in one of the two parental species, and if the allele coming from the same parent is equally higher expressed in the hybrid. A gene is differentially expressed due to *trans*-regulatory changes, if it is differentially expressed in the parental species, but the alleles do not show differential expression in the hybrid

offspring. 'Compensatory' describes the situation, if the gene is not differentially expressed between the parentals but the alleles in the hybrids show differential expression. Figure adapted from (McManus et al., 2010).

Even though it is clear from the literature that both *cis*- and *trans*-regulatory changes contribute to the evolution of phenotypic traits (Hoekstra and Coyne, 2007; Stern and Orgogozo, 2008), genome-wide ASE studies found *cis*-regulatory changes to be more prevalent. This fact is usually explained by the argument that *trans*-regulatory changes would potentially act in a highly pleiotropic manner (Wittkopp et al., 2008). The rational is that mutations in transcription factors or enzymes which are involved in many biological processes, would impact not only one evolving structure but many (He and Zhang, 2006). In contrary, mutations in a *cis*-regulatory region of a gene could have a more tissue-specific function due to the modular nature of the regulatory landscape (reviewed in (Stern and Orgogozo, 2008). Up to now, the question if gene expression divergence results mainly from *cis*- or *trans*-regulatory changes has mostly been tackled by studying adult tissue (e.g. (Coolon et al., 2015; Graze et al., 2009; Wittkopp et al., 2004)). However, since gene expression changes during developmental processes shape adult morphology, it is important to study the mechanisms underlying gene expression divergence also during these early stages.

Even though they are closely related, we found many genes to be differentially expressed between species of the *D. melanogaster* subgroup (Buchberger et al. in prep. (Chapter II), Almudi et al. in prep.) and we used here this model system to study the evolution of gene expression divergence during head and eye development. We combined previous knowledge about regulatory divergence in three species of the *D. melanogaster* subgroup (*D. melanogaster* , *D. simulans* and *D. mauritiana*) with a newly generated ATAC-seq dataset to study, if patterns of *cis*- and *trans*-regulatory divergence can be recapitulated on the basis of open and accessible chromatin regions. Preliminary data surprisingly showed, that gene expression divergence during eye and head development is mainly cause by *trans*-regulatory divergence. Additionally, we describe in Chapter III that the combination of ASE with ATAC-seq datasets indeed allowed to partly recapitulate regulatory divergence by analysing species, stage and tissue specific open chromatin architecture. We revealed, that nucleotide changes in regulatory regions but also their differential accessibility explains parts of the observed *cis*-regulatory divergence. Additionally, not only the coding regions but also the *cis*-regulatroy regions of conserved or *trans*-regulated upstream factors are highly constraint on a sequence level.

3. Chapter I - Transcriptomics supports that pleuropodia of insect embryos function in degradation of the serosal cuticle to enable hatching

The manuscript '**Transcriptomics supports that pleuropodia of insect embryos function in degradation of the serosal cuticle to enable hatching**' is the result of a collaboration with Dr. B. Konopová and Dr. A. Crisp.

The work was conceived and coordinated by Dr. B. Konopová. The manuscript was written by Dr. B. Konopová. I was involved in revising the manuscript.

My contribution for this manuscript includes the following bioinformatic analyses:

- Final cleaning of the transcriptome (filtering, incl. testing for completeness) and blast against Uniprot databases
- Quality assessment of the RNA-seq dataset (quality inspection of raw data, preparation of reads for mapping, principal component analysis (PCA))
- Mapping of RNA-seq reads to the transcriptome
- Differential Expression Analysis
- GO enrichment analysis
- Editing of the draft

I prepared the following figures for the manuscript:

- **Figure 5 A:** Legs and pleuropodia become genetically more different as development progresses
- **Figure 6 A and B:** Dot plot visualization of GO terms enriched in DEGs in the highly secreting pleuropodia

The following figures and tables were summarized and prepared by B. Konopová with the data resulting from my bioinformatics analyses:

- **Figure 5 B:** Legs and pleuropodia become genetically more different as development progresses
- **Figure 7 (only RNA-seq):** Expression profiles of NAGs and CHTs upregulated in the pleuropodia of *Schistocerca* across development

- **Table 1:** Top ten percent of the most abundant transcripts upregulated in the highly secreting pleuropodia of *Schistocerca*
- **Table 2**: RNA-seq differential gene expression of cuticular chitin degrading enzymes in the highly secreting pleuropodia of *Schistocerca*
- **Table 3:** MF proteases that were upregulated in the highly secreting pleuropodia of *Schistocerca*
- **Table 4:** RNA-seq differential gene expression of *Schistocerca* lysozymes in the highly secreting pleuropodia.
- **Table 5:** RNA-seq differential gene expression of *Schistocerca* ecdysone biosynthesis enzymes in the highly secreting pleuropodia.
- **Table S1:** Embryonic transcriptome of *Schistocerca*: numbers of sequenced reads and assembled transcripts.
- **Table S2:** RNA-seq expression analysis: numbers of sequenced and mapped reads.
- **Table S3:** Number of differentially expressed genes at selected levels of stringency.
- **Table S4:** Differential expression of genes, whose expression dynamics in the early stages is known.
- **Table S5 (only RNA-seq):** Comparison between differential expression of selected genes obtained by RNA-seq and real-time RT-PCR.
- **Table S6 - S9:** GO enrichment analyses
- **Table S10 - S15:** transcript annotations
- **Table S16:** RNA-seq differential gene expression of *Schistocerca* ecdysone biosynthesis enzymes in the pleuropodia at diverse stages.
- **Table S17:** *Schistocerca* genes with GO terms "hormone biosynthetic process" upregulated in the highly secreting pleuropodia.

Status of the manuscript:

Published on bioRxiv (doi: http://dx.doi.org/10.1101/584029)

In preparation for submission to Scientific Reports

Title

Transcriptomics supports that pleuropodia of insect embryos function in degradation of the serosal cuticle to enable hatching.

Authors

Barbora Konopová[1,2],*, Elisa Buchberger[2], Alastair Crisp[3]

[1] Department of Zoology, University of Cambridge, United Kingdom

[2] Department of Developmental Biology, University of Göttingen, Germany

[3] MRC Laboratory of Molecular Biology, Cambridge, United Kingdom

* author for correspondence: barbora.konopova@biologie.uni-goettingen.de

Keywords

insect, Orthoptera, RNA-seq, pleuropodia, embryonic organ, gland, moulting fluid, chitinase, immunity, ecdysone

3.1. Abstract

Background

Pleuropodia are limb-derived vesicular organs that transiently appear on the first abdominal segment of embryos from the majority of insect “orders”. They are missing in the model *Drosophila* and little is known about them. Experiments carried out on orthopteran insects eighty years ago indicated that the pleuropodia secrete a “hatching enzyme” that at the end of embryogenesis digests the serosal cuticle to enable the larva to hatch. This hypothesis contradicts the view that insect cuticle is digested by enzymes produced by the tissue that deposited it.

Results

We studied the development of the pleuropodia in embryos of the locust *Schistocerca gregaria* (Orthoptera) using transmission electron microscopy. RNA-seq was applied to generate a comprehensive embryonic reference transcriptome that was used to study genome wide gene expression of 10 stages of pleuropodia development. We show that the mature and secretion releasing pleuropodia are primarily enriched in transcripts associated with transport functions. They express genes encoding enzymes capable of digesting cuticular protein and chitin. These include the potent cuticulo-lytic Chitinase 5, whose transcript rises just before hatching. The pleuropodia are also enriched in transcripts for immunity-related enzymes, including the Toll signaling pathway, melanization cascade and lysozymes.

Conclusions

These data provide transcriptomic evidence that the pleuropodia of orthopterans produce the “hatching enzyme”, whose important component is the Chitinase 5. They also indicate that the organs facilitate epithelial immunity and may function in embryonic immune defence. Based on their gene expression the pleuropodia appear to be an essential part of insect physiology.

3.2. Introduction

An integral part of insect embryogenesis is the transient appearance of enigmatic glandular organs on the first abdominal segment (A1) that are called the pleuropodia (Rathke, 1844; Wheeler, 1890) (Figure 7A-C). These are paired structures that form external vesicles in some species while in others they sink down into the body wall (reviewed in e.g. (Hussey, 1926; Roonwal Mithan Lal and Imms Augustus Daniel, 1936; Wheeler, 1890)). The pleuropodia are peculiarly modified limbs (Bennett et al., 1999; Lewis et al., 2000; Machida, 1981) (Figure 7D,E): their buds emerge in a line with the buds for the walking legs, but unlike the legs, the pleuropodia remain short, the majority of their cells massively enlarge and develop into a transporting-like and secretory epithelium (Bullière, 1970; Louvet, 1975, 1973; Stay, 1977). The pleuropodia degenerate before hatching and are absent in larvae. They have been found in at least some species of nearly all insect "orders" (Figure 7F), but are absent in others, like Diptera, Hymenoptera and advanced Lepidoptera such as silkworms (e.g. (Ando, 1962; Ando and Haga, 1974; Bedford, 1978; Fraulob et al., 2015; Graber, 1889; Hagan, 1931; Heming, 1993; Hussey, 1926; Kamiya and Ando, 1985; Kobayashi et al., 2003; Kobayashi and Ando, 1990; Lambiase et al., 2003; Larink, 1983; Louvet, 1983; Machida, 1981; Machida et al., 2004; Mashimo et al., 2014; Miller, 1940; Miyakawa, 1979; Norling, 1982; Roonwal Mithan Lal and Imms Augustus Daniel, 1936; Rost et al., 2004; S. MILLAM STANLEY and W. GRUNDMANN, 1970; Tanaka et al., 1985; Tsutsumi, 2008; Uchifune and Machida, 2005)). Perhaps because the pleuropodia are missing in the genetic model *Drosophila*, they have been neglected in recent decades. Their function has remained unclear and the genes expressed during their active stages are unknown.

Eighty years ago Eleanor Slifer (H. Slifer, 1937; Slifer, 1938) demonstrated that the pleuropodia of grasshoppers (Orthoptera) are necessary for the digestion of the serosal cuticle (SC) before hatching, to enable the larva to get out of the egg. The SC is a chitin and protein-containing sheet structurally similar to the larval or adult cuticles and is produced by the extraembryonic serosa in early embryogenesis (Goltsev et al., 2009; Jacobs et al., 2015). Shortly before hatching the inner layer of the SC (procuticle) disappears. Slifer (H. Slifer, 1937) showed that when the pleuropodia are removed from fully developed embryos, the SC remains thick and the larva stays arrested in the egg. She proposed that the pleuropodia secrete the "hatching enzyme", a substance likely similar to the cuticle degrading moulting fluid (MF) that is released by the larval epidermis under the old cuticle when the insect is preparing to moult

(Reynolds and Samuels, 1996). The exact molecular composition of this "hatching enzyme" is unknown.

The endocrinologists Novak and Zambre (Novak and Zambre, 1974) argued that this would be an unusual way to digest a cuticle. During larval moulting (Nijhout, 1998) the larval epidermal cells deposit a cuticle and subsequently it is the same epidermal cells, not a special gland that secretes the cuticle degrading MF. Therefore, they proposed that the SC degrading enzymes would most probably be secreted by the serosa itself. They proposed that the pleuropodia instead secrete the moulting hormone "ecdysone", which then stimulates the serosa to secrete the "hatching enzyme". They also suggested that the pleuropodia reach the peak of their activity in very young embryos during katatrepsis when the serosa is still present (Panfilio, 2008).

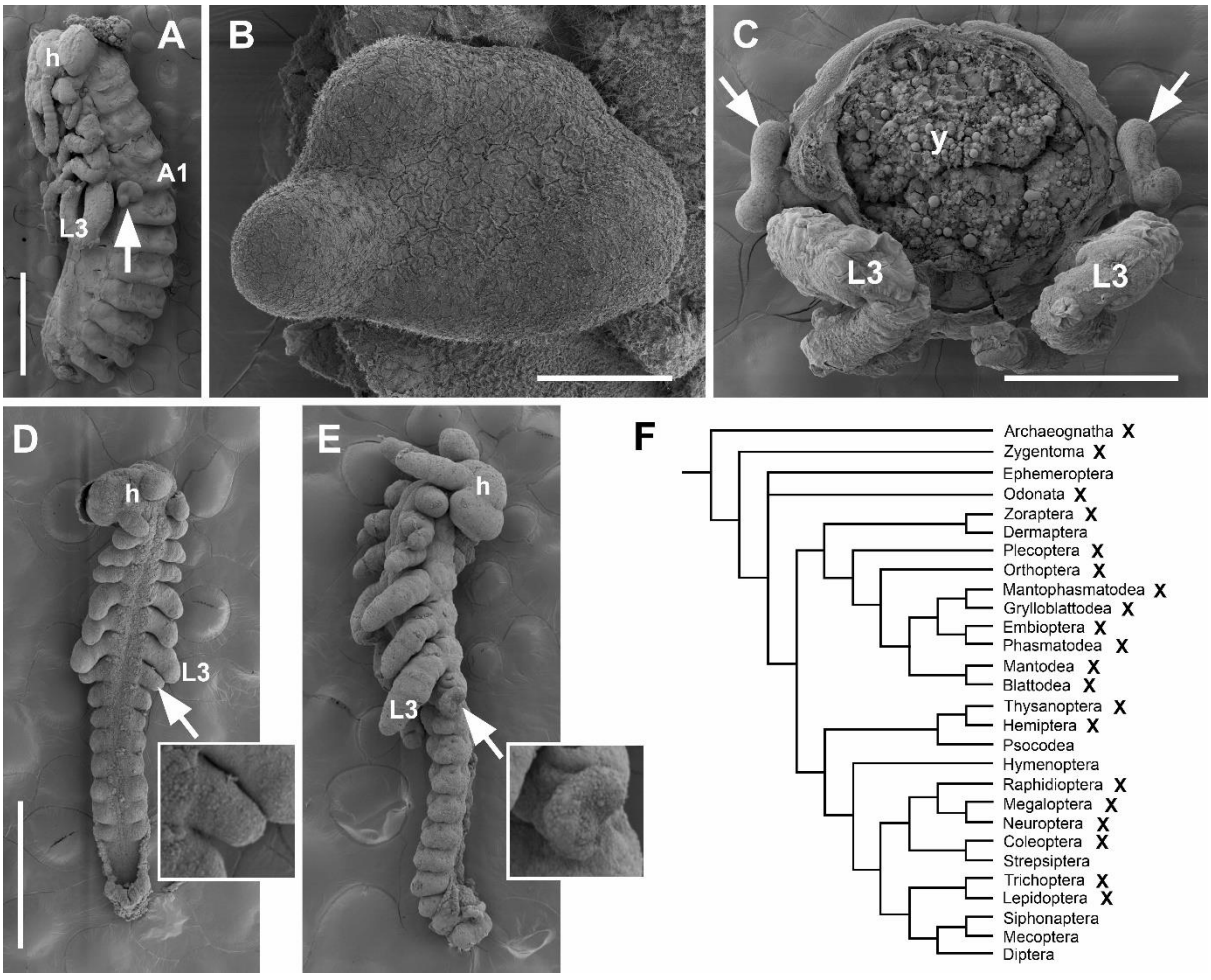


Figure 7. Pleuropodia are limb-derived organs on the first abdominal segment of insect embryos. **A-C.** External morphology of fully developed pleuropodia of *Schistocerca gregaria*. **A.** Embryo before dorsal closure (yolk was removed). **B.** Enlarged left pleuropodium. **C.** Cross section through A1. **D.** and **E.:** Pleuropodia originate by a modification of a limb bud. **D.** Early embryo: all appendages are similarly looking buds. **E.** Older embryo: future legs elongate and the buds on A1 start to take shape of pleuropodia. **F.** Insect phylogenetic tree showing the

presence of pleuropodia among "orders". The cross marks "orders" where at least some species develop pleuropodia. Phylogeny from (Kjer et al., 2016), other references in the text. **A-E** are SEM micrographs. Pleuropodium is marked with an arrow. A1, the first abdominal segment; h, head; L3, hind third. leg; y, yolk. Scale bars: in **A.**, 1 mm; in B., 100 µm; in C.; 500 µm; in **D**., for **D.** and **E.**, 500 µm.

In some insects, including locusts, ultrastructural studies (Bullière, 1970; Louvet, 1975, 1973; Rost et al., 2004; Viscuso and Sottile, 2008) have indeed shown that the pleuropodia secrete granules similar to the "ecdysial droplets" carrying the MF (Locke and Krishnan, 1973). Some of the Slifer's experiments (H. Slifer, 1937) were successfully repeated by others (Jones, 1956) and a substance capable of digesting pieces of SC was even isolated from the pleuropodia (Shutts, 1952). But a proper validation by the state-of-the-art genetic methods that the pleuropodia express genes for enzymes capable to digest the SC is missing.

Here, we identified the mRNAs expressed in the pleuropodia of the locust *Schistocerca gregaria* (Orthoptera). We chose *Schistocerca* as an ideal model, because it has large embryos (eggs over 7 mm) and external pleuropodia that can easily be dissected out, and because the previous experiments testing the function of pleuropodia were carried out in orthopterans. We studied the development of the pleuropodia including using transmission electron microscopy (TEM), and by high-throughput RNA sequencing (RNA-seq) generated transcriptomes from 10 morphologically defined stages. We performed differential gene expression analysis between the pleuropodia and similarly aged hind legs. For mapping of reads we assembled a transcriptome from whole embryos. The goal of this paper was to investigate whether the observed gene expression profile of the pleuropodia is consistent with the idea that these are organs for the secretion of the "hatching enzyme". We show that during their high secretory activity the pleuropodia express genes for cuticle degrading chitinase and proteases that were previously identified in the moulting fluid. This supports the "hatching enzyme hypothesis" (H. Slifer, 1937; Slifer, 1938).

3.3. Results

3.3.1. Development of pleuropodia in the course of *Schistocerca* embryogenesis

Before we could start exploring the genes expressed in the pleuropodia of *Schistocerca* we needed to understand how these organs develop in the locust, i.e. when they are fully differentiated and show activity. Cytological study of developing pleuropodia in grasshopper embryos was previously carried out by Slifer (Slifer, 1938), but the light microscopy that she used does not provide sufficient resolution to distinguish the fine ultrastructure of the cells.

Ultrastructure of pleuropodia by TEM has been described for several insects (Bullière, 1970; Louvet, 1983, 1975, 1973; Rost et al., 2004; Stay, 1977; Viscuso and Sottile, 2008), but a chronological study is missing for *Schistocerca* or any other orthopteran.

Under our conditions *Schistocerca* embryogenesis lasts 14.5 days (100% developmental time, DT) (Figure 8A, S1). We followed the development of the pleuropodia from the age of 4 days (27.6 % DT), when all appendages are similar looking short buds, until just before hatching, day 14 (Figures 8B, S2-3). Simultaneously, we followed the development of the hind leg, which we used for comparison (because pleuropodia are peculiarly modified legs).

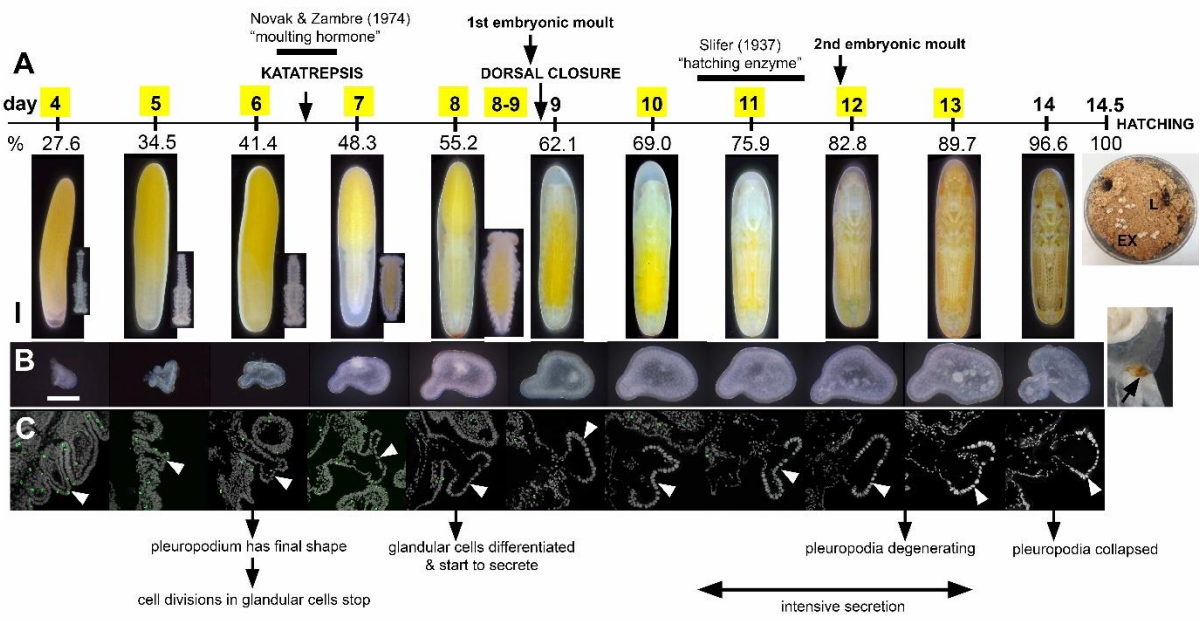


Figure 8. Summary of the development of pleuropodia in *Schistocerca* embryos. **A.** Scheme of *Schistocerca* embryogenesis marking the key developmental events in the embryos and timing of the two experiments on pleuropodia. Numbers above the scale are days from egg-laying, numbers below the scale are percent of embryonic developmental time. Yellow boxes indicate the stages that were sampled for RNA-seq. Eggs with the developing embryos at each stage are shown below the scale, insets for the 4-8 day stages show the embryo dissected out from the egg. **B.** External features of the developing pleuropodia; after hatching part of the stretched exuvia is shown; the degenerated pleuropodium is marked with an arrow. **C.** Paraffin sections through the pleuropodium and surrounding tissue. Pleuropodia are marked with arrowheads. PH3 (green) detects cell divisions in the immature glandular cells (tip of appendage bud) on day 4 and 5, not in later stages. The pleuropodial stalk cells, haemocytes entering the pleuropodia and cells in other tissues were labeled. Nuclei (grey) enlarge from day 6. The text below the pictures refers to the main events in the glandular cells. EX, exuvia; L, larva. Scale bars: in **A.** (eggs), 1 mm; in **B.,** 0.2 mm. Background was cleaned in photos in **A** (see Materials and Methods).

We traced cell divisions in the pleuropodia by using Phosphohistone- 3 as a marker (Figure 8C). The glandular cells were labeled only in the days 4 and 5. From day 6 onwards no cell divisions were detected and the nuclei started to enlarge as the cells became polyploid

(Grellet, 1971). The pleuropodial stalk cells, haemocytes entering the pleuropodia and cells in the other embryonic tissues kept dividing.

Although the pleuropodia get their final external mushroom-like shape just before the embryos undergo katatrepsis (day 6; 41.4% DT) (Figure 8A,B), we found by TEM (Figure 9) that the glandular cells fully differentiate only later, shortly before dorsal closure (day 8; 55.2% DT) (compare the undifferentiated cells in Figure 9F-I, with differentiated cells in Figure 9J-P). At that time these cells form a single-layered transporting-like epithelium (Berridge and Oschman, 1972) and secretion granules inside and outside the cells become visible (Figure 9A-E, J). The granules outside of the cells first appear at the base and in between the long apical microvilli (brush-border) (Figure 9E,J). The whole pleuropodium is covered with a thin embryonic cuticle ("the first embryonic cuticle", EC1); the tips of the microvilli produce fibrous material that is a part of this cuticle (Figure 9E) (compare with similar fibers above the leg epidermis Figure S4).

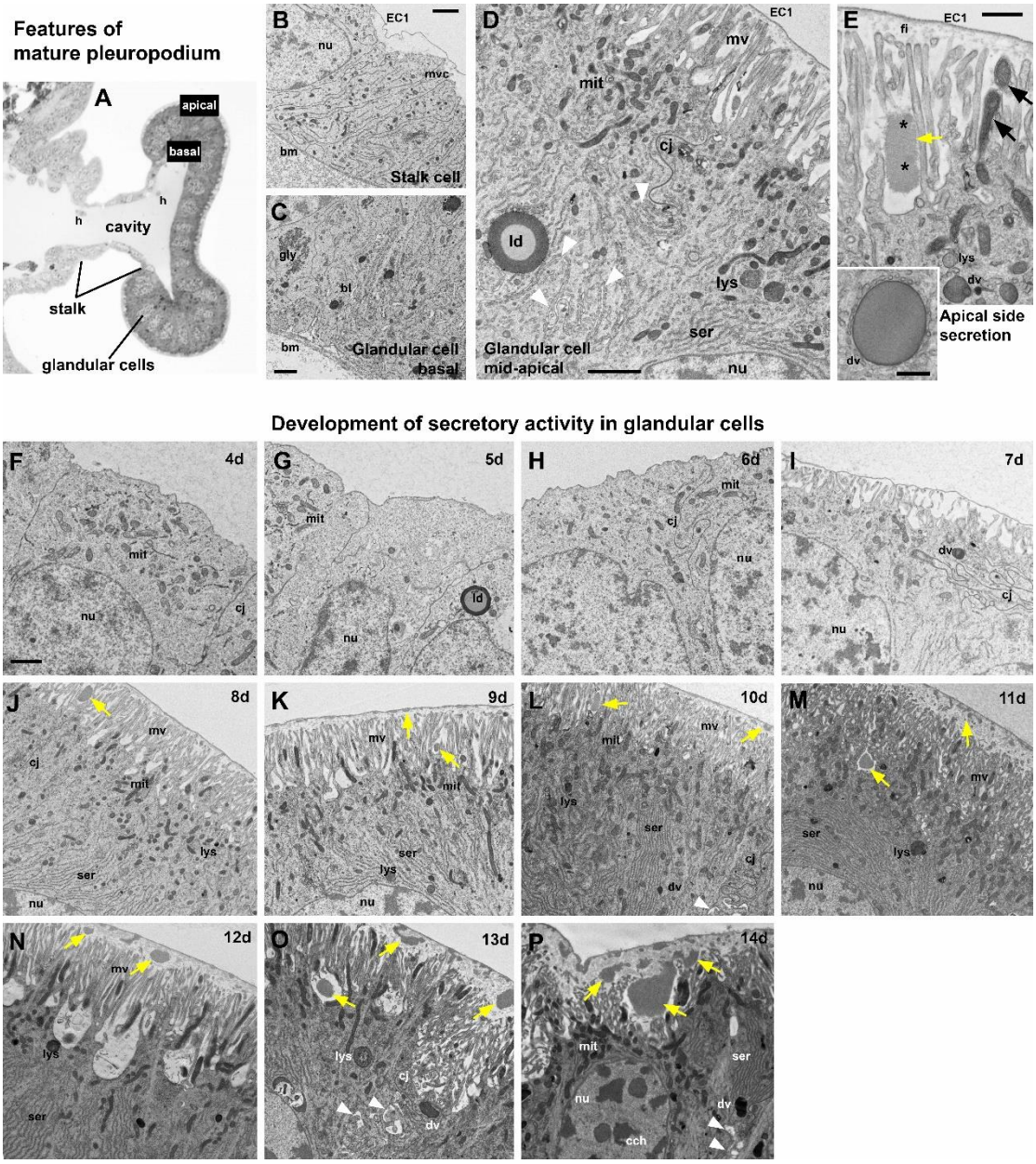


Figure 9. Ultrastructure of the *Schistocerca* pleuropodia. A.-E. Main features of the cells in the fully formed pleuropodia. Pleuropodia just before dorsal closure are shown. **A.** Cross section through the pleuropodium. **B.** Stalk cell. The short microvilli at the apical side are associated with the deposition of fibres in the embryonic cuticle ("the first embryonic cuticle", EC1). **C.-E.** Glandular cells. In **D.** the white arrowheads mark the spaces between neighboring cells. In **E.** the black arrows mark mitochondria inside the microvilli and the asterisks mark spots of different electron-density in the secreted granule. Note that the secretion granule is located at the base of the microvilli (brush-border); the tips of the microvilli produce fibrous material that is a part of the embryonic cuticle EC1. **F.-P.** Ontogenesis of the glandular cells. Note the development of the microvilli (brush border) and the onset of secretion (appearance of secretion granules within and above the microvilli). On day 8 (**J.**) the glandular cells are differentiated, on day 12 (**N.**) patches of the apical side elevate, on day 13 (**O.**) the organelles are disorganized, on day 14 (**P.**) cytoplasm is electron dense (cells shrink), chromatin condensed, but large secretion granules are still present at the base of microvilli and above them. **A.** is a toluidine blue stained semithin section, **B.-P.** TEM micrographs. Secretion granules are marked with yellow arrows. bm, basement membrane; bl, basal labyrinth (infolding of the basal plasma membrane); cj, cell junction; dv, dense vesicle; EC1, the first embryonic cuticle; gly, glycogene; ld, lipid droplet; mit, mitochondria; mv, microvilli; nu, nucleus; ser, smooth endoplasmic reticulum. Scale bars: in **B.**, **C.**, **D.**, **E.** and **F.** for **F.-P.**, 2 µm; inset in **E.**, 500 nm.

As development progresses the secretion granules (inside and outside the cells) become more abundant and are present also above the microvilli (Figure 9K-P). On day 12 the apical side of the glandular cells changes: clusters of microvilli (usually at the borders between cells) elevate (Figure 9N). Later the cells show signs of degeneration, the chromatin condenses and the cell content becomes disorganized (Figure 9O,P). Large secretion granules are still abundant and probably released even on the last day before hatching, when the pleuropodia have shrunk and collapsed (Figures 8B, 9P).

When the embryo moults (apolyses a cuticle and secretes a new one), first at about 8.5 days and again just before 12 days (Figures 8A, S4), ecdysial droplets are present below the apolysed cuticle. These droplets are very similar at both moults (compare Figures S4F and I). They are very similar, but not identical to the granules released by the pleuropodia (Figure 10A,B). The glandular cells of the pleuropodia do not moult and keep the first embryonic cuticle (EC1) their whole life-time.

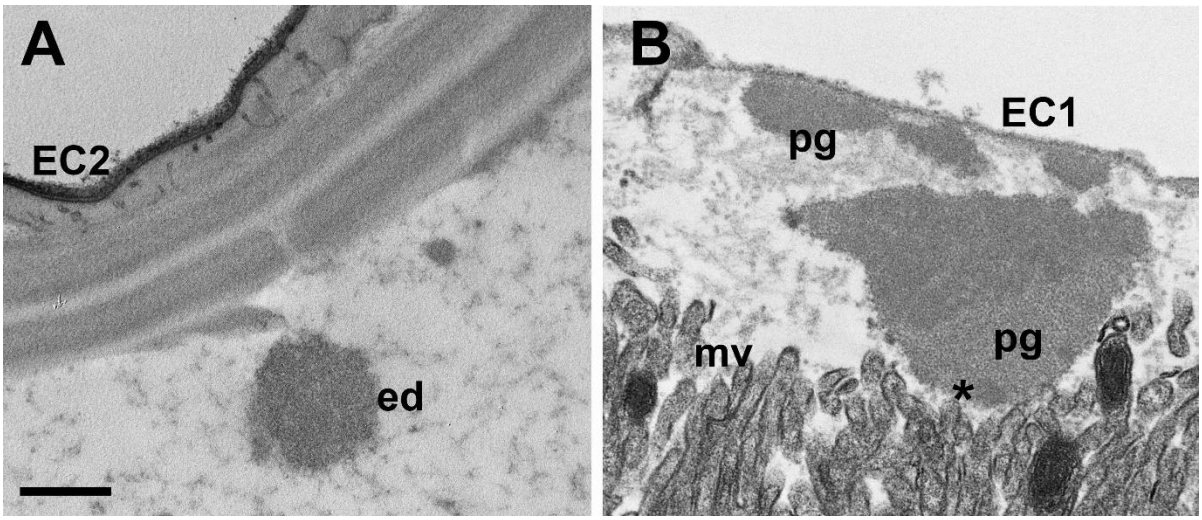


Figure 10. Granules secreted from the pleuropodia resemble ecdysial droplets. A. Ecdysial droplet secreted during the second embryonic moult by hind leg epidermis. **B.** Granules secreted from pleuropodia at the same developmental stage. The pleuropodial granules are typically larger, less compact and with non-homogeneous electrondensity. The "spot" of a different electron-density in the pleuropodial granules is marked with an asterisk. EC1, EC2, the first and second embryonic cuticles; ed, ecdysial droplets; mv, microvilli; pg, granules secreted from the pleuropodia. Scale bar: for **A.** and **B.**, 500 nm.

At hatching, the larva enclosed in the (now apolysed) second embryonic cuticle (EC2) leaves the eggshell and digs through the substrate up to the surface (Bernays, 1971; Konopová and Zrzavý, 2005). Here the EC2 is shed and the degenerated pleuropodia are removed with it ((Roonwal ML and Imms AD, 1936); Figure 8A).

Therefore our observations show that the timing of the high secretory activity corresponds to the stages when Slifer (H. Slifer, 1937) demonstrated the presence of the "hatching enzyme" (Figure 8A). Next we looked at what genes are expressed in the pleuropodia at this time.

3.3.2. Generation of a comparative RNA-seq dataset from developing pleuropodia and legs of *Schistocerca*

To find out what genes are upregulated in the pleuropodia of *Schistocerca*, we applied a comparative genome wide expression analysis using RNA-seq. We generated a comprehensive embryonic transcriptome (see details in Materials and Methods) that served as reference for the analysis. This transcriptome consists of 20 834 transcripts (Table S1). Its completeness was assessed using the open-source software BUSCO (Simão et al., 2015; Waterhouse et al., 2017). 95.6%, 96.3% and 94.6% of the Metazoa, Arthropoda and Insecta orthologs, respectively, were found, a level comparable to published "complete" transcriptomes.

To gain insights into the gene expression dynamics of pleuropodia development, we dissected pleuropodia from 10 embryonic stages and isolated their mRNAs. In parallel, we dissected hind legs for the same 10 stages to generate a comparative transcriptomic dataset. In total we sequenced pairs of samples (pleuropodia and legs) from 10 developmental stages and performed a differential expression analysis between legs and pleuropodia for each stage (Figure 8A, Table S2). A principal component analysis (PCA) confirmed that legs and pleuropodia are not only morphologically very similar at early stages, but share a common transcriptomic landscape as well (Figure 11A). The number of differentially expressed genes (DEGs) rises as development progresses (Figure 11B, Table S3).

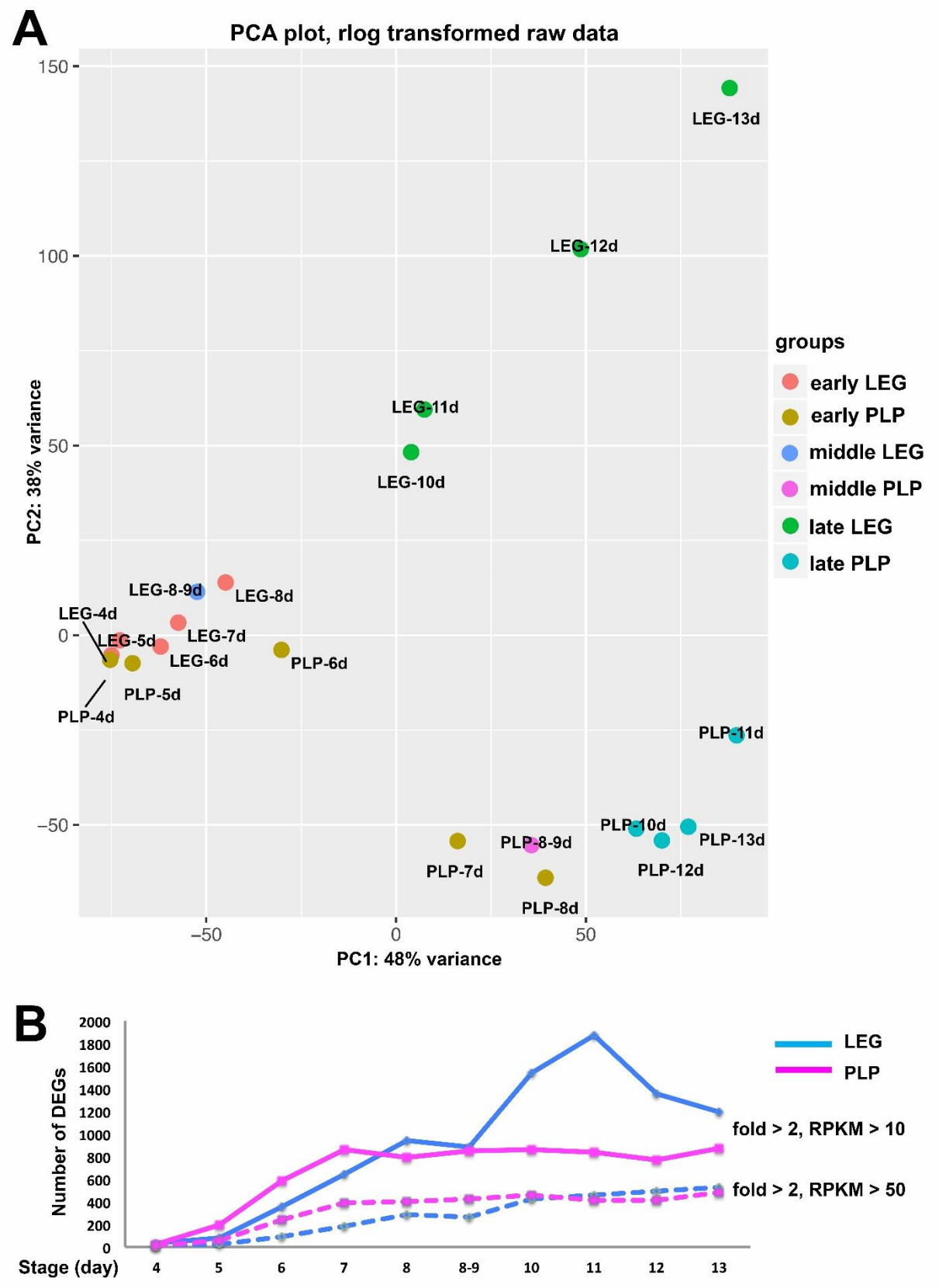


Figure 11 Legs and pleuropodia become genetically more different as development progresses. A. PCA on genes expressed in legs and pleuropodia at 10 embryonic stages (rlog transformed read counts). The expression profile diversifies with development, consistent with the observation that the two tissues develop into two different structures (starting from day 6). Samples from young embryos are genetically more similar and cluster together, while samples from advanced stages are genetically more distant and also separated on the plot. **B.** Number of

DEGs at two levels of stringency (RPKM ≥ 10 and fold change ≥ 2 was considered as a threshold for a gene to be differentially expressed). LEG, DEGs downregulated in pleuropodia and upregulated in legs, PLP, DEGs upregulated in pleuropodia and downregulated in legs.

For several genes whose expression dynamics in the pleuropodia were already known, such as *Ubx*, *abd-A*, *dll* and *dac* (Angelini et al., 2005; Bennett et al., 1999; Hughes and Kaufman, 2002; Prpic et al., 2001; Tear et al., 1990; Zhang et al., 2005), we confirmed that they were up- or downregulated in our RNA-seq data as predicted (Table S4). To further validate the RNA-seq dataset, we carried out real-time RT-PCR on 46 selected genes in several stages (in total in 176 cases) and got results consistent with the sequencing data (Table S5). Therefore, we are confident that we can identify important factors that are relevant for pleuropodia function and development.

3.3.3. Identification of genes upregulated in the intensively secreting pleuropodia

Since we wanted to focus specifically on the pleuropodia with high secretory activity we pooled the data from the samples 10, 11 and 12 days together, separately for pleuropodia and legs, and treated them as triplicates. These three samples cover the stages from the embryos after the dorsal closure, when the pleuropodia intensively release secretion granules, but are not in advanced state of degeneration (day 13) (Figures 10A, 9L-N). We performed differential expression analysis and gene ontology (GO) enrichment analysis with genes upregulated in legs and pleuropodia. We identified 781 transcripts upregulated in the pleuropodia (compared to the legs) and 1535 downregulated (Table S3). Table 1 shows the top 10% of the most highly abundant transcripts (measured in RPKM units, "reads per kilobase of transcript per million reads mapped") that we found upregulated in the pleuropodia.

Table 1. Top ten percent of the most abundant transcripts upregulated in the highly secreting pleuropodia of *Schistocerca*.

Transcript ID	Protein	Characteristics	Immunity[a]	Cuticle digestion[b]	RPKM		Fold change
					legs	pleuropodia	
SgreTa0017702	x				23.07	15186.05	658.36
SgreTa0007897	C-type lysozyme	anti-bacterial protein	x		42.93	14452.15	336.64
SgreTa0002988	Uncharacterized, contains DUF4773 domain				15.16	9112.05	601.19
SgreTa0005052	x				13.37	7950.98	594.48
SgreTa0001636	Serine protease	proteolysis	x	x	49.38	7578.31	153.48
SgreTa0008851	Chitin binding Peritrophin-A	perotrophic matrix protein			9.12	6836.42	749.88
SgreTa0017707	I-type lysozyme	anti-bacterial protein	x		12.20	6712.31	550.26
SgreTa0007042	x				7.04	6650.18	944.25
SgreTa0004599	Alpha-tocopherol transfer protein	intermembrane lipid transfer			8.99	5848.12	650.71
SgreTa0009217	x				5.03	5384.56	1070.14
SgreTa0003175	Collagen				32.25	5220.96	161.87
SgreTa0007886	Alpha-N-acetylgalactosaminidase	carbohydrate catabolism			3.85	4372.63	1134.69
SgreTa0002109	x				2.20	3016.31	1372.07
SgreTa0017715	Serine protease, Snake-like	proteolysis, Toll signaling	x	x	70.55	2947.46	41.78
SgreTa0017664	Chitinase 5	cuticular chitin degradation		x	79.32	2620.11	33.03
SgreTa0002467	Neutral endopeptidase 24.11	proteolysis		x	62.26	2282.01	36.66

SgreTa0004397	x				11.21	2266.30	202.21
SgreTa0002828	x				1.77	2188.14	1234.00
SgreTa0006539	Serpin, 88E-like	serine protease inhibitor	x		32.42	2152.14	66.38
SgreTa0001321	Glycosyl hydrolase, Myrosinase 1-like	carbohydrate catabolism			3.93	2070.40	527.16
SgreTb0011177	x				1.38	1884.79	1369.32
SgreTa0008335	x				54.24	1812.38	33.41
SgreTa0003635	Alpha-tocopherol transfer protein	intermembrane lipid transfer			2.23	1800.68	806.99
SgreTb0003860	Serine protease, H2-like	proteolysis	x	x	77.42	1727.41	22.31
SgreTa0013418	x				0.87	1484.98	1710.66
SgreTa0014009	Angiotensin-converting enzyme	proteolysis		x	65.76	1457.47	22.16
SgreTa0006966	Pro-phenol oxidase subunit 2	immunity, melanization	x		144.78	1347.43	9.31
SgreTa0000425	6-phosphofructo-2-kinase	glycolysis			93.52	1346.50	14.40
SgreTa0003661	Serine protease, Easter-like	proteolysis	x	x	29.50	1332.79	45.18
SgreTa0006960	Glutamate dehydrogenase mitochondrial	nitrogen and glutamate metabolism			172.56	1327.45	7.69
SgreTa0017670	Xaa-Pro aminopeptidase	proteolysis		x	2.89	1322.01	457.96
SgreTb0000759	Cathepsin L	proteolysis, lysosomal enzyme		x	105.63	1308.36	12.39
SgreTa0014684	x				1.30	1294.87	994.80
SgreTa0007025	Insect pheromone-binding protein A10/OS-D	chemoreception			1.77	1224.20	692.95
SgreTa0006282	Cytochrome P450 CYP4G102	synthesis of hydrocarbons, anti-dehydration			2.91	1196.27	410.93
SgreTa0009515	Sensory neuron membrane protein, 1-like	chemoreception			3.33	1188.81	357.50
SgreTa0008528	C-type lysozyme	anti-bacterial protein	x		8.61	1159.55	134.71
SgreTa0009095	Catalase	redox homeostasis	x		355.15	1158.27	3.26
SgreTb0039135	x				3.53	1119.22	316.71
SgreTa0001486	Lipopolysaccharide-induced tumor necrosis factor-alpha factor homolog	lysosomal degradation			45.83	1109.33	24.20
SgreTb0039012	x				14.29	1060.82	74.25
SgreTa0009747	Serpin (27-like)	serine protease inhibitor, melanization	x		14.49	1054.67	72.80
SgreTa0013400	Peroxiredoxin, 5-llke	redox homeostasis	x		101.10	1034.15	10.23
SgreTa0017395	x				5.08	1004.86	197.64
SgreTa0017712	x				15.59	990.41	63.53
SgreTa0005600	Beta-N-acetylglucosaminidase NAG2	cuticular chitin degradation		x	15.10	939.60	62.21
SgreTa0000783	Serine protease, Snake-like	proteolysis	x	x	4.30	917.47	213.59
SgreTa0006651	Uncharacterized, contains Transcription activator MBF2 domain				1.62	907.98	561.49
SgreTa0017657	Putative serine protease, K12H4.7-like / Serine carboxypeptidase	proteolysis		x	2.31	904.26	391.60
SgreTa0017700	Peroxidase	redox homeostasis	x		5.36	874.51	163.25
SgreTa0002600	Uncharacterized, contains DUF3421 domain				0.97	870.73	894.35
SgreTb0019827	Tob	antiproliferative protein, transcription corepressor			141.26	846.86	5.99
SgreTa0017854	x				0.85	838.89	981.74
SgreTa0007774	Lysosomal-associated membrane protein	lysosomal membrane protein			185.20	822.81	4.44
SgreTa0015156	x				27.45	804.82	29.32
SgreTa0007809	Tetraspanin	scaffolding protein in cell membrane			63.04	799.76	12.69
SgreTa0004471	Leucine rich repeat	membrane glycoprotein			74.88	797.35	10.65
SgreTa0004278	Fatty acyl-CoA reductase, waterproof-like	lipid metabolism			1.75	733.39	417.99
SgreTa0014626	V-type proton ATPase proteolipid subunit	proton transporting ATPase			190.76	708.56	3.71
SgreTa0016256	Bax inhibitor 1	negative regulation of apoptosis and autophagy			237.58	692.52	2.91
SgreTa0001469	Sodium/potassium-transporting ATPase subunit alpha	sodium:potassium exchanging ATPase			119.60	685.51	5.73
SgreTa0007426	Serine protease, Easter-like	proteolysis	x	x	0.66	673.43	1023.60
SgreTa0007081	Vigilin	RNA binding, sterol metabolism			247.46	655.61	2.65
SgreTa0013328	Ferritin	iron ion transport, iron sequestration	x		238.10	651.31	2.74
SgreTa0002155	Uncharacterized serine protease inhibitor	serine protease inhibitor	x		33.83	646.73	19.12
SgreTa0014303	x				176.21	645.78	3.66
SgreTa0017577	Aquaporin	water channel			0.39	635.34	1638.96
SgreTa0013377	Phosphoenolpyruvate carboxykinase [GTP]	gluconeogenesis			13.56	628.95	46.37
SgreTa0005752	Alpha-tocopherol transfer protein	intermembrane lipid transfer			12.98	594.56	45.79
SgreTa0014098	Phospholipase B-like	lipid degradation			206.76	577.99	2.80
SgreTa0000856	Transposase-like				25.93	576.67	22.24
SgreTa0008861	x				0.37	541.63	1456.67
SgreTa0017826	Sodium:neurotransmitter symporter	solute:sodium symport			0.49	540.53	1104.10
SgreTb0019287	x				3.11	528.47	169.79
SgreTa0015520	Protein yellow	melanization	x		2.75	520.09	189.08
SgreTb0006243	I-type lysozyme	anti-bacterial protein	x		16.96	519.35	30.62
SgreTa0009559	Gram-negative bacteria binding protein 3	pathogen recognition	x		15.40	510.04	33.13

[a] proteins related to immune response

[b] proteins that participate in larval moulting; some of them are known, other anticipated to digest cuticular chitin and protein (e.g., present in the MF)

For the sake of clarity we summarized redundant GO terms in representative GO-groups (Figure 12; the full set of enriched GO terms are presented in Tables S6,S7; GOs enriched at each developmental stage separately are in Tables S8,S9) (see Materials and Methods). Our results show that the genes downregulated in the pleuropodia (upregulated in the legs) are enriched in GO terms associated with development and function of muscle tissue, cell division and DNA synthesis. This is in agreement with our and previous observations that the pleuropodia lack muscles, while at these stages the legs are differentiating, developing muscles and their cells are still dividing (Figure 8C). The pleuropodia downregulate genes for the development of mesoderm, which is consistent with the morphological observation that they are formed by ectodermal cells (Figure 9A).

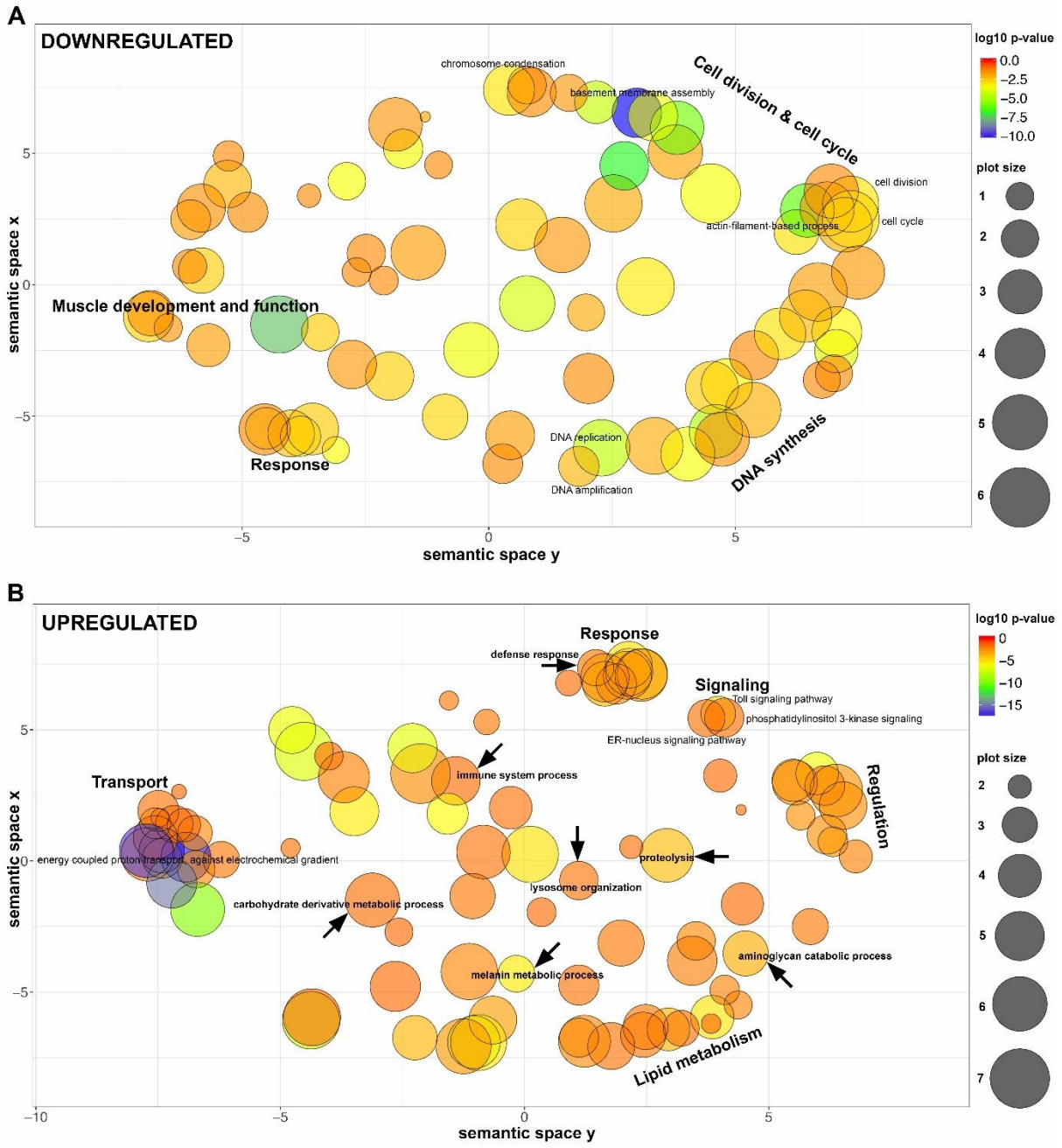


Figure 12. Dot plot visualization of GO terms enriched in differentially expressed genes in highly secreting pleuropodia. Representative groups of GO terms enriched in genes that are **A.** downregulated in pleuropodia (in comparison to legs) and **B.** upregulated in pleuropodia. Major clusters are labeled. Relevant GOs are marked with an arrow. Bubble color indicates the p-value of the GO term, the size indicates the frequency of the GO term in the underlying Gene Ontology Annotation (GOA) database (bubbles of more general terms are larger).

The upregulated genes are primarily enriched in GO terms (Figure 12, Table S7) associated with transport thus genetically confirming the morphological observations that the pleuropodia are transporting organs. These include genes for transporters present in typical insect transporting epithelia (Chintapalli et al., 2013), such as the energy providing V-ATPase and Na^+, K^+ ATPase (Table S10). We found enriched GO terms linked with lysosome organization, consistent with the observation that the pleuropodia contain numerous lysosomes (Figure 9,

(Louvet, 1975)). We also found a large cluster of GO terms associated with lipid metabolism, consistent with the abundant smooth endoplasmic reticulum in the cells. Therefore, the pool of genes expressed in the pleuropodia is in agreement with the morphology of the organs. Among the novel findings are upregulation of genes associated with immunity, as well as with carbohydrate derivative metabolism, aminoglycan catabolic process and proteolysis: these might contain genes for degradation of the SC. Next we looked at selected genes in a detail.

3.3.4. The pleuropodia upregulate genes for cuticular chitin degrading enzymes

Insect cuticle is digested by a cocktail of chitin and protein degrading enzymes (Reynolds and Samuels, 1996; Zhang et al., 2014). Cuticular chitin is hydrolyzed by a two-enzyme system composed of a β-N-acetyl-hexosaminidase (NAG) and a chitinase (CHT) (Zhu et al., 2007). Both types of enzymes, a NAG and a chitinase, have to be simultaneously present for efficient hydrolysis of chitin (Fukamizo and Kramer, 1985). Previous studies have shown that only particular NAGs and CHTs are capable of efficiently digesting the type of chitin present in the insect cuticle (see below).

Insect NAGs were classified into 4 major classes, of which chitinolytic activity was demonstrated for group I and II (Table 2) (Hogenkamp et al., 2008; Rong et al., 2013). Our transcriptome contains 4 NAG transcripts, each representing one group (Table 2, Figures 13A-D, S5A, S6A). All were upregulated in the pleuropodia. Among them the *Sg-nag2* for the chitinolitic NAG group II had the highest expression (among 46 most highly "expressed" genes, Table 1) and fold change between legs and pleuropodia. The abundance of transcripts for the chitinolitic NAGs starts to rise from day 6 (Figure 13A, B) when the glandular cells in the pleuropodia begin to differentiate morphologically (Figs 7, 9). The expression profile of *Sg-nag2*, that we have chosen for validation, was similar by RNA-seq and real-time RT-PCR (compare Figure 13B and B').

Table 2. RNA-seq differential gene expression of cuticular chitin degrading enzymes in highly secreting pleuropodia of *Schistocerca*.

Family	Group	Protein	*Schistocerca* gene	UP/DOWN[a]	Fold change	Expression[b]
ß-N-acetylhexosaminidase	I	NAG1	*Sg-nag1*	**UP**	7.85	124 (15.88%)
	II	NAG2	*Sg-nag2*	**UP**	62.21	46 (5.89%)
	III	Fused lobes	*Sg-fdl*	**UP**	14.18	592 (75.8%)

	IV	Hex	*Sg-hex*	**UP**	47.37	306 (39.18%)
chitinase-like	I-Major "moulting" chitinases	Chitinase 5	*Sg-cht5-1*	**UP**	33.03	15 (1.92%)
			Sg-cht5-2	**UP**	234.78	400 (51.21%)
	II-"Moulting" chitinases	Chitinase 10	*Sg-cht10-1*	na[c]		
			Sg-cht10-2	ns[d]		
	III-Cuticle assembly chitinases	Chitinase 7	*Sg-cht7-1*	ns		
			Sg-cht7-1	ns		
			Sg-cht7-1	ns		
	IV-Gut, fat body and other chitinases	Chitinase 8	*Sg-cht8-1*	na		
			Sg-cht8-1	na		
			Sg-cht8-1	na		
		Chitinase 6	*Sg-cht6-1*	ns		
			Sg-cht6-2	ns		
		Chitinase 2	*Sg-cht2*	**UP**	2.81	188 (24.07%)
	V-Imaginal disc growth factors	IDGF	*Sg-idgf-1*	**UP**	20.97	391 (50.06%)
			Sg-idgf-2	ns		
			Sg-idgf-3	ns		

[a] upregulated (UP)/ downregulated (DOWN)

[b] the DEGs were ranked according to their RPKM (in descending order), the number describes the position of the DEG in the ranked table; top 25% highlighted in black, others in descending level of grey

[c] not applicable (expression low to undetectable in both samples, transcript filtered out)

[d] *not significant*

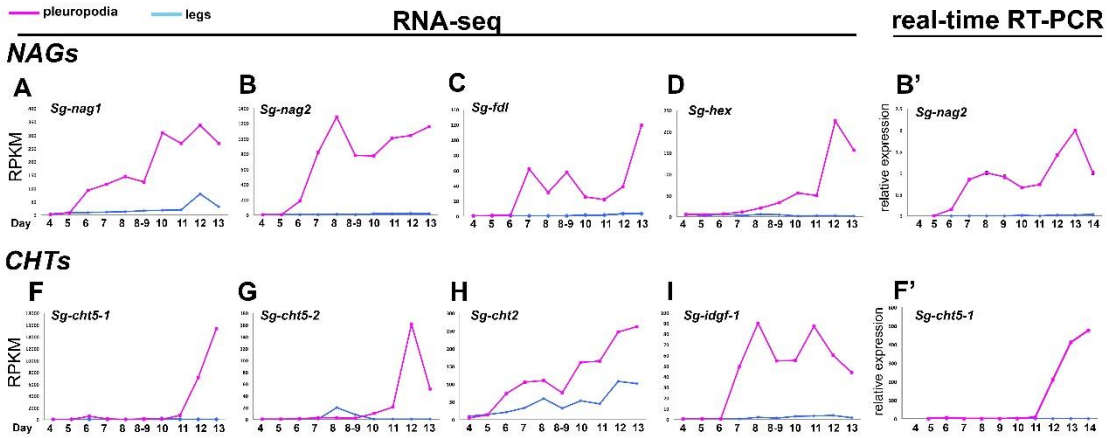


Figure 13. **Expression profiles of NAGs and CHTs upregulated in the pleuropodia of *Schistocerca* across development.** Top row: NAGs, bottom row: CHTs. **A-D.** and **F-I.** RNA-seq, Expression in single-sample sequencing is shown. **B'.** and **F'.** real-time RT-PCR. **B'.** is the same gene as in **B.** and **F'**. is the same gene as in **F.** Analysis of 3-4 technical replicates is shown. Expression in day 8 was set as 1.

To see if the pleuropodia are the major source of the *Sg-nag2* transcript in the embryo, we looked at its expression in various parts of the body (head, thorax, abdomen with pleuropodia, abdomen from which pleuropodia were removed) using real-time RT-PCR (Figure 14A,B). We performed this analysis in embryos on day 6, when the pleuropodia are still immature, day 8, just at the onset of the secretory activity, day 10 and day 12 during active secretion. During all of the stages the abdomen with pleuropodia had the highest expression (A+ in Figure 14B), although the expression was lower in the youngest sample (day 6) compared to the samples from older embryos (day 8, 10 and 12). This shows that the pleuropodia are the major source of mRNAs for this cuticle-degrading NAG.

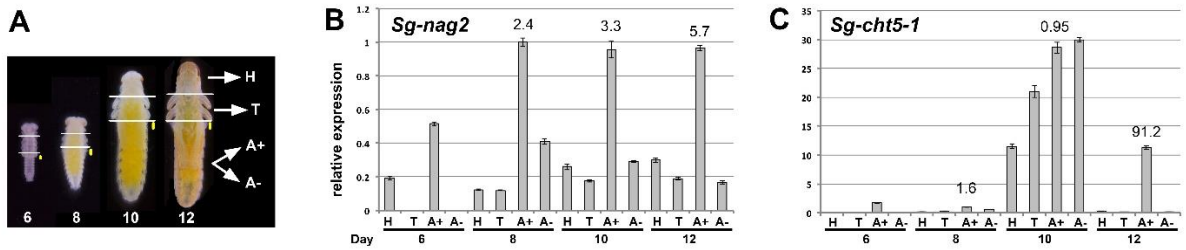


Figure 14. Real-time RT-PCR expression analysis of Sg-nag2 and Sg-cht5-1 on cDNA from parts of Schistocerca embryos. A. cDNA was prepared from mRNAs isolated from parts of embryos at the age of 6, 8, 10 and 12 days: H, head; T, thorax; A+, abdomen with pleuropodia; A-, abdomen without pleuropodia. For each age the same number of body parts was used (5-10) and RNA was resuspended in the same volume of water. The size of the pleuropodium is indicated by the yellow dot. **B.** and **C.** expression of Sg-nag2 and Sg-cht5-1, respectively. Analysis of 3-4 technical replicates is shown. Expression in A+8 (abdomen with pleuropodia at stage when the organs first become differentiated) was set as 1. Numbers above A+ expression is fold change from A- of the same age.

The insect CHTs have been classified into several groups (Noh et al., 2018; Zhu et al., 2016), of which the major role in the digestion of cuticular chitin is played by Chitinase 5 and (perhaps with a secondary importance) by Chitinase 10 (Qu et al., 2014; Zhu et al., 2008) (Table 2; the classification of CHTs into 5 major groups that we use here is based on (Zhu et al., 2008)). Some chitinases, for example, are expressed in the gut, trachea and fat body, where they are likely involved in digestion of dietary chitin, turnover of peritrophic matrix and immunity, other chitinases organize assembly of new cuticle (Merzendorfer, 2013; Noh et al., 2018; Pesch et al., 2016).

Our transcriptome contains 16 full or partial transcripts of CHTs representing all of the major CHT groups (Table 2, Figure S5B, S6B). The pleuropodia specifically upregulate both of the genes for Chitinase 5, homologs of *cht5-1* and *cht5-2* from the locust *Locusta migratoria* (Li et al., 2015). One of the transcripts, *Sg-cht5-1*, was among the top 15 most highly expressed genes (Table 1). The predicted amino acid sequence contains a conserved catalytic domain and

a signal peptide, and thus is likely to be active and secreted, respectively (Figure S5B). The other upregulated CHTs were homologs of *Cht2* and *Idgf*. By contrast, the *Schistocerca* homolog of cht-10 that also has a role in cuticular chitin hydrolysis and required for larval moulting (Pesch et al., 2016; Zhu et al., 2008) had low expression in both legs and pleuropodia.

We next focused on the transcript of the major chitinase, *Sg-cht5-1*. Unlike the NAGs, both RNA-seq and real-time RT-PCR have shown that the expression of this CHT is low in the early secreting stages, rises only later around day 12 and reaches highest levels when the pleuropodia are already degenerating (day 13 and 14) (Figure 13 F,G,F'). Also real-time RT-PCR on cut embryos has shown that the pleuropodia are a major source of the *Sg-cht5-1* mRNA on day 12 but not before (the high expression in the whole embryo on day 10 could be linked to the second embryonic moult and was also observed with *Sg-cht7*, although not with *Sg-cht10*, Figure S8). These data show that the pluropodia before hatching express a cuticle-degrading chitinase.

3.3.5. Pleuropodia upregulate transcripts for some proteases that could digest a cuticle

Our GO enrichment analysis has shown that the secreting pleuropodia are enriched in transcripts for genes associated with proteolysis (Figure 12, Table S11). Transcripts for proteases and their inhibitors are abundant among the top 10 per cent of the most highly "expressed" upregulated DEGs (Table 1). To see if the upregulated transcripts encode enzymes that are associated with digestion of insect cuticle, we compared our data with the enzymes identified in the complete proteomic analysis of the MF from the lepidopteran *Bombyx mori* (Liu et al., 2018; Zhang et al., 2014). Out of 69 genes that we searched, we found homologs or very similar genes in *Schistocerca* transcriptome for half of them (35). This made in total 75 transcripts, of which 27 were upregulated (7 among the top 10 per cent most highly expressed) and 15 downregulated (Table 3, S12). The prominent MF protease Carboxypeptidase A (Sui et al., 2009; Zhang et al., 2014) and the Trypsin-like serine protease known to function in locust moulting (Wei et al., 2007) were not upregulated in the pleuropodia. These data indicate that the pleuropodia upregulate transcripts for proteolytic enzymes associated with the degradation of the cuticle and would be able to contribute to digest the SC.

Table 3. MF proteases that were upregulated in the highly secreting pluropodia of *Schistocerca*.

MF protein[a]	Blast query[b]	*Schistocerca* transcript ID[c]	homolog/similar[d]	RPKM PLP	Fold change UP

Putative peptidase	D2KMR2	SgreTa0000627	similar	131.75	3.14
Aminopeptidase N-12	I3VR83	SgreTb0018983	similar	35.86	4.35
Neutral endopeptidase 24.11	Q9BLH1	**SgreTa0002467**	similar	2282.01	36.66
	Q9BLH1	SgreTa0017692	similar	133.30	240.28
	Q9BLH1	SgreTb0039123	similar	219.35	186.96
Ecdysteroid-inducible angiotensin-converting enzyme	Q9NDS8	**SgreTa0014009**	similar	1457.47	22.16
	Q9NDS8	SgreTa0017728	similar	62.71	57.08
Carboxypeptidase E-like	H9IST0	SgreTa0000925	homolog	139.81	10.95
Angiotensin-converting enzyme-like	H9IZ41	SgreTa0003298	homolog	23.64	5.65
Aminopeptidase N-like	H9JEW9	SgreTa0017219	homolog	391.03	437.93
Digestive cysteine protease 1, cathepsin L	H9JHZ1	SgreTa0000627	homolog	131.75	3.14
Serine carboxypeptidase	H9J242	**SgreTa0017657**	homolog	904.26	391.60
Serine protease HP21 precursor	H9JJA9	SgreTa0017649	similar	179.69	24.45
Trypsin-like serine protease - fibroin heavy chain	H9JPA8	**SgreTa0001636**	homolog	7578.31	153.48
Serine protease, Easter-like	Q2VG86	SgreTa0003188	homolog	485.97	837.45
	Q2VG86	**SgreTa0003661**	homolog	1332.79	45.18
	Q2VG86	SgreTa0006780	homolog	103.37	14.76
	Q2VG86	SgreTa0007424	homolog	29.62	79.13
	Q2VG86	SgreTa0007425	homolog	123.69	72.31
	Q2VG86	SgreTb0037249	homolog	21.76	249.74
	Q2VG86	SgreTb0039879	homolog	305.63	544.04
	H9JLZ4	SgreTa0010219	similar	46.12	20.75
	H9JLZ4	SgreTb0039024	similar	11.71	22.11
Serine protease 1	H9JXY6	**SgreTb0003860**	homolog	1727.41	22.31

Serine protease, Snake-like	H9IWW2	**SgreTa0000783**	similar	917.47	213.59

[a] proteomic sequencing of MF of the lepidopteran *Bombyx mori* (Zhang et al., 2014; Liu et al., 2018)

[b] Uniprot ID for blast on *Schistocerca* transcriptome

[c] transcripts in bold were among the top 10% most highly "expressed" upregulated DEGs (Table 1)

[d] *considered as homologous, if reciprocal blast retrieved the query sequence*

3.3.6. Pleuropodia are enriched in transcripts for immunity-related proteins

An observation that was not anticipated was the upregulation of genes for proteins involved in immunity (Buchon et al., 2014; Lemaitre and Hoffmann, 2007) (Figures 12, 15, Table S13). This is especially interesting, because immunity related proteins have been found in the MF (Zhang et al., 2014). It is in agreement with that the cells in the pleuropodia are a type of barrier epithelium (Bergman et al., 2017; Buchon et al., 2014; Lemaitre and Hoffmann, 2007), which enables the contact between the organism and its environment. Barrier epithelia (e.g., the gut, Malpighian tubules or tracheae) constitutively express genes for immune defense.

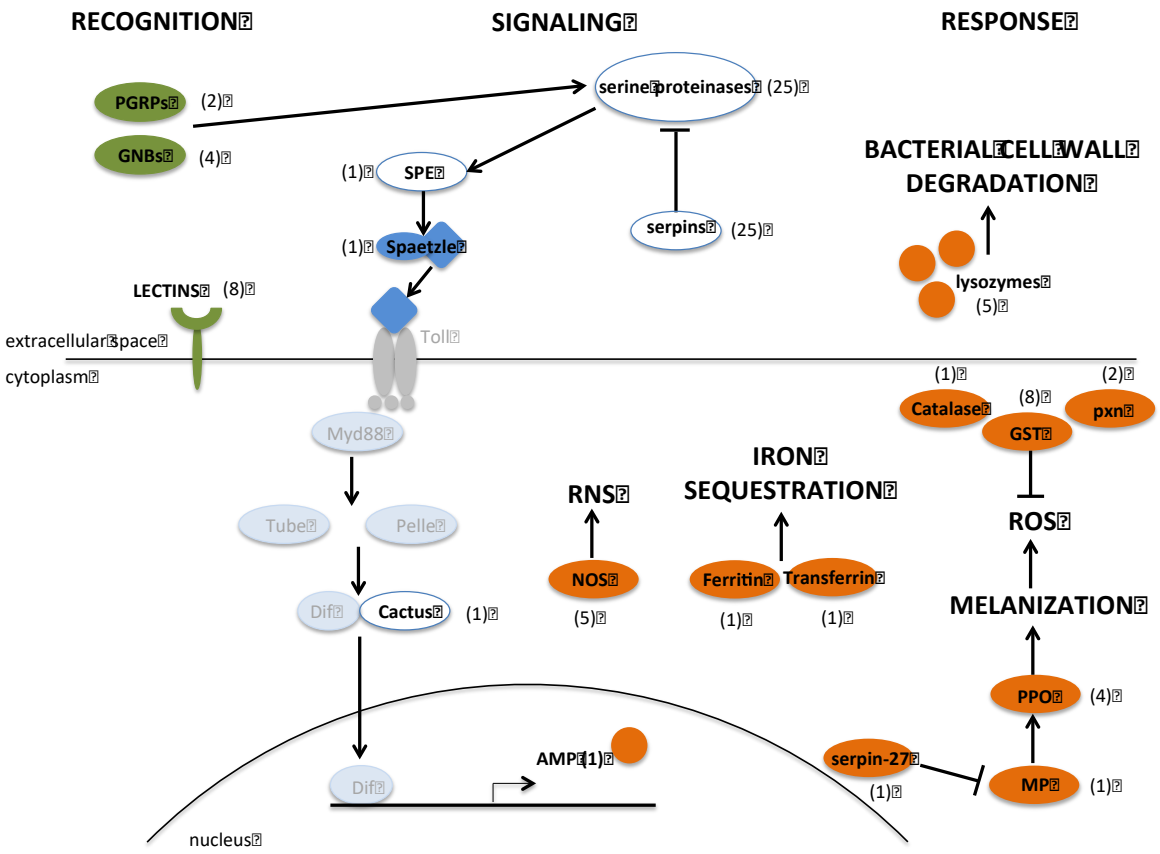


Figure 15. Schematic representation of the key immunity-related genes expressed in the highly secreting pleuropodia of *Schistocerca*. Proteins whose transcripts were found in the pleuropodia are in black, number in the brackets is the number of upregulated transcripts. Proteins whose transcripts were not upregulated are in

grey. Out of the total 25 serine proteases and 25 serpins, 14 and 15 are known to function in Toll signaling, respectively. AMP, antimicrobial peptide; GNBP, gram-negative bacteria-binding protein; GST, glutathione S-transferase; MP, melanization protease; NOS, nitric oxide synthase; PGRP, peptidoglycan recognition preotein; PPO, pro-phenoloxidase; pxn, peroxiredoxin; RNS, reactive nitrogen species; ROS, reactive oxygen species; SPE, Spaetzle-processing enzyme.

In total we found upregulated 99 transcripts (13 per cent of the upregulated genes) for immunity-related proteins. These include proteins at all three levels, the pathogen recognition, signaling and response (Figure 15, Table S13). From the four signaling pathways, Toll was upregulated, but not IMD or JAK/STAT, and from the JNK signaling we found c-Jun. Genes for a range of immune responses were upregulated, including production of reactive nitrogen species (RNS), melanization, genes for lysozymes and one antimicrobial peptide (AMP) similar to Diptericin.

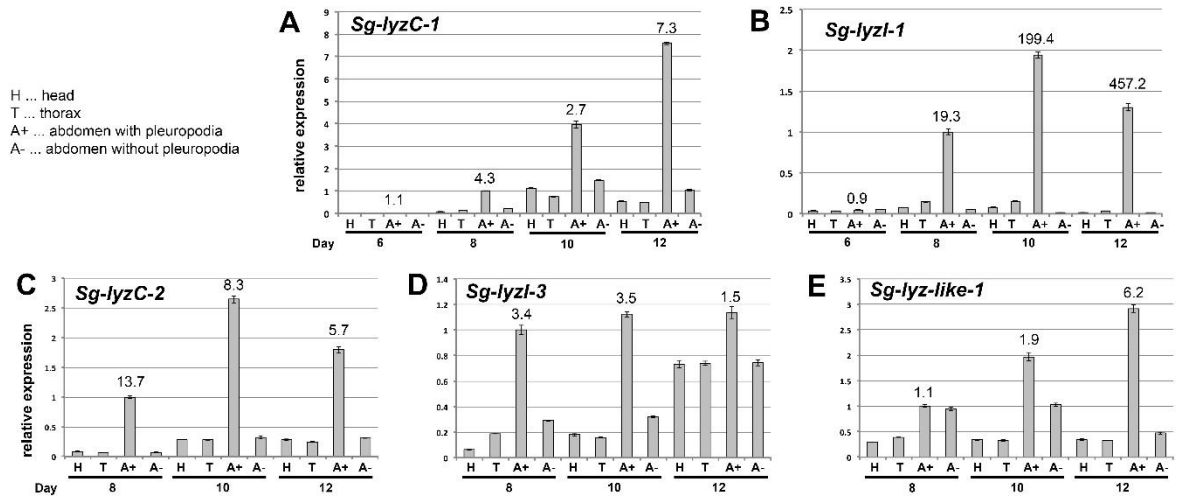


Figure 16. Real-time RT-PCR expression analysis of genes for lysozymes on cDNA from parts of *Schistocerca* embryos. cDNA was prepared from mRNAs isolated from parts of embryos at the age of 6, 8, 10 and 12 days. For each age the same number of body parts was used (5-10) and RNA was resuspended in the same volume of water. Analysis of 3-4 technical replicates is shown. Expression in A+8 (abdomen with pleuropodia at stage when the organs first become differentiated) was set as 1. Numbers above A+ expression is fold change from A- of the same age.

The transcripts for lysozymes were among the most highly expressed (Table 1) and we chose to focus on them. Lysozymes are secreted proteins that kill bacteria by breaking down their cell wall. Our *Schistocerca* transcriptome contains 9 genes for lysozymes, 7 of which were upregulated (Table 4, Table S14). The second most highly expressed DEG was a transcript for a C-type lysozyme *(SgLyzC-1)* that was previously shown to have anti-bacterial properties in *Schistocerca* (Mohamed et al., 2016) (Table 1). We examined expression of 5 selected genes on cut embryos by real-time RT-PCR (Figure 15). Our data showed that the pleuropodia are the major source of mRNAs for these genes.

Table 4. RNA-seq differential gene expression of *Schistocerca* lysozymes in the highly secreting pleuropodia.

Lysozyme type	Gene	UP/DOWN[a]	Fold change	Expression[b]
C-type lysozyme	*SgLyzC-1*	**UP**	336.64	2 (0.26%)
	SgLyzC-2	**UP**	134.71	37 (4.74%)
I-type lysozyme	*SgLyzI-1*	**UP**	550.26	7 (0.90%)
	SgLyzI-2	ns[c]		
	SgLyzI-3	**UP**	30.62	76 (9.73%)
	SgLyzI-4	DOWN	-34.41	1251 (81.50%)
	SgLyzI-5	ns		
Lysozyme-like	*SgLyz-like-1*	**UP**	192.68	150 (19.21%)
	SgLyz-like-2	ns		

[a] upregulated (UP)/ downregulated (DOWN)

[b] the DEGs were ranked according to their RPKM (in descending order), the number describes the position of the DEG in the ranked table; shading as in Table 2

[c] *not significant*

3.3.7. The pleuropodia do not upregulate the pathway for ecdysone biosynthesis

Previous work has suggested that pleuropodia may be embryonic organs producing the moulting hormone ecdysone (Novak and Zambre, 1974). During post-embryonic stages, ecdysone is synthesized in the prothoracic glands and several other tissues by a common set of enzymes (Niwa and Niwa, 2014; Ou et al., 2016), some which have been characterized in the locusts (Lenaerts et al., 2016; Marchal et al., 2012, 2011; Sugahara et al., 2017). As shown in *Drosophila*, these genes are expressed in diverse cell types in embryos, and when the larval prothoracic glands are formed their expression co-localizes there (Chávez et al., 2000; Niwa et al., 2004; Petryk et al., 2003; Warren et al., 2004, 2002).

Out of the nine genes critical for ecdysone biosynthesis, only one (*dib*) was upregulated in the highly secreting pleuropodia (Table 5, S15). One gene (*spo*) was downregulated. The pleuropodia are not enriched in the whole pathway at any time of development, including around katatrepsis, in which experiments supporting the synthesis of moulting hormone were carried out (Table S9, S16). Under the GO term “hormone biosynthetic process” enriched in the

intensively secreting pleuropodia (Table S7, S17) we found a gene *Npc2*a that encodes a transporter of sterols including precursors of ecdysone. It is also required for ecdysone biosynthesis, but indirectly and in the cells it functions as a general regulator of sterol homeostasis (Huang et al., 2007). We conclude that our transcriptomic data provide little evidence that the pleuropodia are involved in ecdysone biosynthesis.

3.4. Discussion

3.4.1. Pleuropodia of *Schistocerca* express genes for the "hatching enzyme"

The first demonstration of the physiological role of the pleuropodia comes from the experiments carried out on a grasshopper *Melanoplus* (closely related to *Schistocerca*), by Eleanor Slifer (H. Slifer, 1937). When she took embryos before hatching (Figure 8) and separated anterior and posterior halves by ligation the SC was digested only in the part of the egg with the pleuropodia. Surgical removal of the pleuropodia prevented SC digestion in the whole egg. Slifer's hypothesis that the pleuropodia secrete the "hatching enzyme" was criticized by Novak and Zambre (Novak and Zambre, 1974): if the deposition and digestion of the SC is similar to the cuticle turnover during larval moulting, then the "hatching enzyme" is produced by the serosa. They believed that the pleuropodia reach the peak of their activity in embryos during katatrepsis (45% development) and participate on SC digestion indirectly by secreting ecdysone to stimulate the serosa.

Our ultrastructural observations on staged pleuropodia of *Schistocerca* have shown that the glandular cells only begin to differentiate just at the time of katatrepsis (45% DT) and do not secrete at that time. This would explain why no digestive effect on the SC was detected by Novak and Zambre (Novak and Zambre, 1974) using a homogenate from *Schistocerca* pleuropodia isolated at this stage. The release of granular secretion starts just before the dorsal closure (55% DT) and intensifies before hatching. This is in agreement with previous observations on some stages of the pleuropodia in other orthopterans (Louvet, 1975; Viscuso and Sottile, 2008).

Our RNA-seq analysis revealed that the secreting pleuropodia highly express genes encoding enzymes that are capable of digesting a typical chitin-protein insect cuticle. These include genes for proteolytic enzymes similar to those present in the moulting fluid and cuticular chitin-degrading NAGs and Chitinase 5. The pleuropodia also express genes for Chitinase 2 and Idgf, which have low effect on cuticular chitin digestion, but were shown to

organize proteins and chitin fibres during cuticle deposition (Pesch et al., 2016). These CHTs may organize the fibres in the cuticle secreted by the pleuropodia (Figure 9).

In combination with RT-PCR we showed that, while the expression of the *Sg-nag1* and *Sg-nag2* started to rise in parallel with the differentiation of the glandular cells, the *Sg-cht5-1* and *Sg-cht5-2* transcripts raised shortly before hatching. Chitinase 5 is a critical chitin-degrading chitinase in insects: it is highly abundant in the moulting fluid and its silencing in diverse insects including locusts leads to failure in larval moulting (Li et al., 2015; Pesch et al., 2016; Xi et al., 2015; Zhang et al., 2014; Zhu et al., 2008). Our data indicate that the sudden rise in the expression of *cht5* in the pleuropodia at the end of embryogenesis and presumably secretion of this CHT into the extraembryonic space is the key component of the "hatching enzyme" effect (H. Slifer, 1937; Slifer, 1938) in locusts and grasshoppers.

3.4.2. Pleuropodia in some other insects could secrete the "hatching enzyme" and their function may also vary among species

There is evidence to suggest that the process occurs similarly in some insect. As in orthopterans, the pleuropodia of the rhagophthalmid beetle *Rhagophthalmus ohbai* release secretion soon after katatrepsis and SC rapidly degrades just shortly before hatching (Kobayashi et al., 2003). In the large water true bugs from the family Belostomatidae, the male carries a batch of eggs on his back. It is believed that the detachment of the eggs just before hatching is also caused by the secretion from the pleuropodia (Tanizawa et al., 2007).

The molecular mechanism of SC degradation may also vary between insects and as previously hypothesized (Novak and Zambre, 1974) the serosa may also contribute to the SC degradation. The serosa of the beetle *Tribolium*, expresses *cht10* and *cht7* (Jacobs et al., 2015), of which the former CHT is important for cuticular chitin digestion. Silencing of *cht10*, but not *cht5* prevented larvae from hatching (Zhu et al., 2008). Transcripts for *cht10* were not upregulated in the pleuropodia of *Schistocerca*. This suggests that the SC is degraded by enzymes produced by both, the serosa and the pleuropodia and that the indispensable roles in cuticle digestion are played by different enzymes in different insects.

In some insects the pleuropodia may not be involved in hatching but have another function. In the viviparous cockroach *Diploptera punctata* (Stay, 1977), the secretion from the pleuropodia is very low and the large pleuropodia of the melolonthid beetle *Rhizotrogus majalis* have not been observed to produce any secretion granules at all (Louvet, 1983). In dragonflies,

one of the more basal lineages of insects, the secretion likely has a different function than in orthopterans, because their SC is not digested before hatching (Andō, 1962). The special epithelium in the pleuropodia shares features with transporting epithelia (Louvet, 1973; Stay, 1977) that function in water transport and ion balance (Berridge and Oschman, 1972). Our data do not exclude this function, but it is yet to be tested.

3.4.3. The pleuropodia of *Schistocerca* are enriched in transcripts for enzymes functioning in immunity

We found that many of the genes expressed in the pleuropodia encode proteins involved in immunity (Lemaitre and Hoffmann, 2007). This indicates that the pleuropodia are also organs of epithelial immunity, similar to other barrier epithelia in postembryonic stages (such as the gut) (Bergman et al., 2017), which are in a constant contact with microorganisms. The pleuropodia differ from such tissues in that they are not directly exposed to the environment, but enclosed in the eggshell, seemingly limiting their contact with microorganisms. Proteins associated with immune defense are also found in the MF (Zhang et al., 2014), where they prevent invasion of pathogens through a "naked" epidermis after the separation of the old cuticle from the epidermis in the process of apolysis. As found in the beetle *Tribolium*, during the early embryonic stages the frontier epithelium providing the egg with an immune defense (Jacobs et al., 2014) is the extraembryonic serosa. The serosa starts to degenerate after katatrepsis and disappears at dorsal closure (Panfilio, 2008). The pleuropodia of *Schistocerca* differentiate just before dorsal closure, suggesting that they take over this defense function in late embryogenesis. It will be interesting to clarify in the upcoming research whether apart from their role in hatching the pleuropodia are important organs for fighting against potential pathogens that have gained access to the space between the embryo and the eggshell.

3.4.4. Conclusions

The pleuropodia of *Schistocerca* have morphological markers of high secretory activity in the second half of embryogenesis after the definitive dorsal closure is finished. Transcriptomic profiling indicate that the conclusions that Eleanor Slifer drew from her experiments over eighty years ago that the pleuropodia secrete cuticle degrading enzymes, were correct. The pleuropodia likely have other functions, such as in immunity. The pleuropodia are specialized embryonic organs and an important though neglected part of insect physiology.

3.5. Material and Methods

3.5.1. Insects

Schistocerca gregaria (gregarious phase) were obtained from a long-term, partly inbred colony at the Department of Zoology, University of Cambridge. Eggs were collected into aluminium pots filled with damp sand. The pots were picked up after 2 (most samples) or 4 hours and incubated at 30°C.

3.5.2. Description of embryonic stages

Embryos and appendages were dissected in phosphate buffer saline (PBS). Whole eggs were bleached in 50 per cent household bleach to dissolve the chorion. All were photographed in water or PBS using the Leica M125 stereomicroscope equipped with DFC495 camera and associated software. Photos were processed using Adobe Photoshop CC 2017.1.1. Photos of eggs and embryos that illustrate the stage (Figure 8A and S1) had the background cleaned using the software (removal of the tools that hold the photographed objects in place).

3.5.3. Immunohistochemistry on paraffin sections

Embryos were dissected in PBS and pieces including posterior thorax and anterior abdomen (older embryos) or mid thorax plus whole abdomen (young embryos) were fixed in PEMFA (4% formaldehyde in PEM buffer: 100 mM PIPES, 2.0 mM EGTA, 1.0 mM $MgSO_4$) at room temperature (RT) for 15-30 minutes, then washed in PBT (PBS with 0.1 % Triton-X 100) and stored in ethanol at -20°C.

Samples were cleared in 3x10 minutes in Histosol (National Diagnostics) at RT, infiltrated with paraffin at 60°C for 2-3 days, embedded in moulds and hardened at RT. Sections 6-8 µm thick were prepared on a Leica RM2125RTF microtome. The slides with sections were washed with Histosol, ethanol, then step wise re-hydrated to PBT. Incubations were carried out in a humidified chamber. Slides were blocked with 10% sheep serum (Sigma-Aldrich) in PBT for 30 minutes at RT, incubated with Phospho-Histone H3 antibody (Invitrogen) diluted with PBT 1:130 at 4°C overnight, washed and incubated with Alexa Fluor 568 anti-rabbit secondary antibody (Invitrogen) diluted 1:300 at RT for 2 hours, washed and incubated with DAPI (Invitrogen) diluted 1:1000. Sections were imaged with a Leica TCS SP5 confocal microscope and photos processed using Fiji (https://fiji.sc).

3.5.4. Transmission (TEM) and scanning (SEM) electron microscopy

For TEM embryos were removed from the chorion in PBS and pieces of posterior thorax to anterior abdomen were fixed in 2.5-3.0% glutaraldehyde in 0.1 M phosphate buffer pH7.2 for a few hours at room temperature and then at 4°C for several days. Each pleuropodium and leg were then separated and embedded into 2 % agar. Small cubes of agar with the tissue were incubated in osmium ferrocyanide solution (3 % potassium ferricyanide in cacodylate buffer with 4 mM calcium chloride) for 1-2 days at 4°C , then in thiocarbohydrazide solution (0.1 mg thiocarbohydrazide from Sigma-Aldrich, and 10 ml deionized water dissolved at 60°C) and protected from light for 20-30 minutes at RT, then in 2% aqueous osmium tetroxide 30-45 minutes at RT and in 1% uranyl acetate (maleate buffered to pH 5.5) at 4°C overnight. Washing between each step was done with deionized water. Samples were dehydrated in ethanol, washed with dry acetone, dry acetonitrile, infiltrated with Quetol 651 resin (Agar Scientific) for 4-6 days and hardened in moulds at 60°C for 2-3 days. Semithin sections were stained with toluidine blue. Ultrathin sections were examined in the Tecnai G280 microscope.

For SEM whole embryos were dissected out of the chorion in PBS, fixed in 3% glutaraldehyde in phosphate buffer similarly as above. They were post-fixed with osmium tetroxide, dehydrated through the ethanol series, critical point dried, gold coated, and observed in a FEI/Philips XL30 FEGSEM microscope. Photos from TEM and SEM were processed using Adobe Photoshop CC 2017.1.1.

3.5.5. Preparation of the reference transcriptome

Whole embryo transcriptome: Eggs from each 1-day egg collection incubated for the desired time were briefly treated with 50% bleach, washed in distilled water and frozen in liquid nitrogen. Total RNA was isolated with TRIzol reagent (Invitrogen), treated with TURBO DNase (Invitrogen) and purified on a column supplied with the RNAeasy Kit (Quiagen). The purified RNA from each day (14 samples) was pooled into 4 samples: day 1-4, 5-7, 8-10 and 11-14. 10 µg of RNA from each of the 4 samples was sent to BGI (Hong Kong). The total RNA was enriched in mRNA by using the oligo(dT) magnetic beads and cDNA library was prepared. 100 bp paired-end (PE) reads were sequenced on Illumina HiSeq 2000; numbers of the reads obtained are in Table S2. Non-clean reads were filtered using filter_fq software (removes reads with adaptors, reads with unknown nucleotides larger than 5% and low quality reads). Transcripts from all samples were assembled separately using the Trinity software (release 20130225) (Grabherr et

al., 2011) with parameters: --seqType fq --min_contig_length 100; --min_glue 4 --group_pairs_distance 250; --path_reinforcement_distance 95 --min_kmer_cov 4. Transcriptes from the 4 assemblies were then merged together to form a single set of non-redundant transcripts using TGICL software (v2.1) (Pertea et al., 2003) with parameters: -l 40 -c 10 -v 20.

Legs and pleuropodia transcriptome (age about 8.5-8.75 days): The appendages were dissected in cold RNase-free PBS (treated with diethyl pyrocarbonate) and total was RNA isolated and cleaned as described above. 10 µg of RNA from each leg sample and pleuropodium sample were transported to the Eastern Sequence and Informatics Hub (EASIH), Cambridge (UK). cDNA libraries were prepared including mRNA enrichment. 75 bp PE reads were sequenced on Illumina GAIIX; numbers of the reads obtained are in Table S2. The reads were trimmed to the longest contiguous read segment for which the quality score at each base is greater than a Phred quality score of Q = 13 (or 0.05 probability of error) using the program DynamicTrim (v. 1.7) from the package SolexQA ((Cox et al., 2010) http://solexaqa.sourceforge.net/). The trimmed reads were then filtered to remove sequence adapter using the program cutadapt (v. 0.9; http://code.google.com/p/cutadapt/). Sequences shorter than 40 base pairs were discarded. Trimmed reads were used to de novo assemble the transcriptome using Velvet (v. 1.1.07; (Zerbino and Birney, 2008); http://www.ebi.ac.uk/~zerbino/velvet/) (commands: -shortPaired –fastq; -short2 –fastq; -read_trkg yes) and Oases (v. 0.2.01; (Schulz et al., 2012); http://www.ebi.ac.uk/~zerbino/oases/) (commands: -ins_length 350). Velvet is primarily used for de-novo genome assembly; here, the contigs that were output by Velvet were used by the complementary software package Oases to build likely transcripts from the RNA-seq dataset. K-mer sizes of 21, 25 and 31 were attempted for the two separate samples as well as the combined samples and optimal K-mer sizes of 21 were found for both samples.

Transcripts for the reference transcriptome were selected from the embryonic and legs and pleuropodia transcriptome. The transcripts were first merged with evigene ((Gilbert, 2013) version 2013.03.11) using default parameters. Because this selection of transcripts eliminated some genes (gene represented by zero transcripts, although the transcripts were present in the original transcriptomes), we repeated the step with less strict parameters (cd-hit-est - version 4.6, with -c 0.80 -n 5). This second selection contained several genes represented by more transcripts, thus we aligned selection 1 and 2 to each other to identify, which genes in selection

1 were missing. Selection 1 was then completed with the help of selection 2 by adding the missing transcripts. The quality and completeness of the resulting transcriptome was assessed and edited in the following steps. First, we removed several redundant transcripts manually: these were found by blasting diverse insect sequences (queries) against the *Schistocerca* transcriptome using the local ViroBLAST interface (Deng et al., 2007). Some transcripts were edited manually, such as when we found that two transcripts were combined into one, resulting in an alignment against two protein sequences (*Schistocerca* transcript blasted against NCBI database) we split the respective transcripts. Second, we blasted the whole transcriptome against itself and removed redundant sequences, if the alignment was spanning at least 300bp with a sequence identity of at least 98% (Blast+ suite, version 2.6.0) (Camacho et al., 2009). The longer transcript was kept in all cases. Transcripts shorter than 200 bp were discarded. All these steps were carried out in R (R Development Core Team, 2008) and sequences were handled using the Biostrings package (Pagès et al., 2017).

3.5.6. Sequence analysis

Basic transcript analysis was done by CLC Sequence Viewer7 (QIAGEN). Signal peptide and transmembrane regions were predicted by Phobius (Käll et al., 2007); http://phobius.binf.ku.dk/index.html). To annotate the newly assembled transcriptome, the freely available annotation pipeline Trinotate (version 3.1.1) was used (Haas et al., 2013). The longest candidate ORF of each sequence was identified with the help of the inbuilt TransDecoder (Haas et al., 2013); https://github.com/TransDecoder/TransDecoder/wiki) software.

A blast was run against Uniprot sequences specific for *Schistocerca gregaria, Locusta migratoria, Apis melifera, Tribolium castaneum, Bombyx mori* and *Drosophila melanogaster* (blastx with default parameter and -max_target_seqs 1) and against nr database using Blast2GO (Götz et al., 2008).

3.5.7. RNA-seq expression analysis

Pleuropodia and hind legs from embryos at the same age (day 4, 5, 6, 7, 8, 10, 11, 12 and 13) were dissected in cold RNase-free PBS and total RNA was isolated as described for samples for the reference transcriptome, but cleaned with RNA Clean & Concentrator (Zymo Research). 1 µg of RNA from each sample was sent to BGI (Hong Kong). The mRNA enrichment

and cDNAs preparation as described above. 50 bp single-end (SE) reads were sequenced on Illumina HiSeq 2000. Over 45 million reads were sequenced from each sample (Table S2).

A pair of samples from mixed embryos 8-9 days that was used for the preparation of the reference transcriptome (described above) was also included in the expression analysis, but prior to mapping, the 75bp PE reads were trimmed to 50 bp, using Trimmomatic in the paired-end mode (version 0.36) using the CROP function (CROP:50) (Bolger et al., 2014). A single pleuropodium or leg sample was sequenced from each stage.

The quality of the sequenced reads was assessed with the help of the FastQC software. All samples consistently showed a Per base sequence quality of >30. Reads were mapped to the Reference transcriptome with Bowtie2 (version 2.2.5) using default parameter and the –local alignment mode (Langmead et al., 2009). The trimmed pairs of reads were concatenated for each stage and treated as single reads. A PCA plot was generated to assess if differences in sequencing type and processing (SE samples and PE samples day 8-9) had an effect, which was not the case. This plot was prepared by using the plotPCA() function in the DESeq2 R package (Love et al., 2014); the count matrix was transformed with the rlog() function. The R package HTSFilter (Rau et al., 2013) was used with default parameters to filter constantly low expressed genes and 12988 transcripts were left.

The differential expression analysis was performed with the NOISeq R package (2.22.1; (Tarazona et al., 2011). Reads were first normalized using the RPKM method (Mortazavi et al., 2008). We used NOISeq-sim to find the differentially expressed genes between legs and pleuropodium for each stage with the following parameters: k = NULL, norm ="n", pnr =0.2, nss =5, v = 0.02, lc=1, replicates ="no", following the recommendations by the authors for simulation of "technical replicates" prior to differential expression analysis without replicates. Additionally differentially expressed genes between active pleuropodia and legs at the same stage were assessed (treating samples from day 10, 11 and 12 as replicates) using the NOISeq-real algorithm with the following parameters: k=0.5, norm="n", factor="type", nss=0, lc=1, replicates = "technical". To define significantly, differentially expressed genes, the probability ("prob") threshold was set at 0.7 for single stage comparisons and 0.8 for the triplicated comparison, RPKM ≥ 10 and fold change ≥ 2 for both single stage and triplicated comparisons (based on the expression of the genes whose expression dynamics in the pleuropodia were already known, Table S4).

3.5.8. GO enrichment

The transcriptome was blasted against the whole UniProt/Swiss-Prot database to assess the corresponding GO terms. Only blast hits with an e-value <= 1e-5 were considered for the subsequent GO annotation. GO enrichment of differentially expressed genes was performed using the R package GOSeq (version 1:30.0, (Noh et al., 2018) implemented in the Trinotate pipeline (see above). Enriched GO-terms were summarized and visualized with REVIGO (Supek et al., 2011). Dot plots were prepared from DEGs selected at thresholds: RPKM>50, fold change >3.

3.5.9. Real-time RT-PCR

Tissues were dissected, total RNA was isolated and DNase treated the same way as for sequencing and cleaned with RNA Clean & Concentrator (Zymo Research). cDNA was synthesized with oligo-dT primer (Invitrogen) 0.5 µg (legs, pleuropodia) or 1 µg (pieces of embryos) of the RNA using ThermoScript RT-PCR System (Invitrogen) at 55°C. The cDNA was diluted to concentration 40 ng/µl and 5 µl was used in a reaction containing 10 µl of SYBR Green PCR Master Mix (Applied Biosystems) and 5 µl of a 1:1 mix of forward and reverse primers (each 20nM in this mix). Reactions were run in the LightCycler480 (Roche) and analyzed using the associated software (release 1.5.0 SP1) according to the comparative Ct method and normalized to the *eEF1α* gene. Primers (Table S18) were designed with Primer3PLUS program (Untergasser et al., 2007). To check for the presence of a single PCR product, the melting curve was examined after each run and for each pair of primers at least 2 finished runs were visualized on a 2 % agarose gel.

The program was: denaturation: 95°C for 10 minutes (1 cycle), amplification: 95°C for 10 seconds, 60°C for 15 seconds, 72°C for 12 seconds (40 cycles) melting: 95°C for 5 seconds, 60°C for 1 minute, 95°C.

3.6. List of abbreviations

CHT: chitinase, DEG: differentially expressed gene; EC1, EC2: the first and the second embryonic cuticle, respectively; GO: gene ontology; LEG: hind leg(s); MF: moulting fluid; NAG: β-N-acetyl-hexosaminidase; PCA: principal component analysis; PLP: pleuropodium (pleuropodia); RPKM: reads per kilobase of transcript per million reads mapped; SC: serosal cuticle

3.7. Data availability

The sequencing data generated and analyzed during the study are available in the NCBI repository, BioProject ID PRJNA524786 (the reference transcriptome has the accession number GHHP00000000, the version described in this paper is the first version, GHHP01000000).

3.8. Competing interests

The authors declare that they have no competing interests.

3.9. Funding

This work was supported by Human Frontier Science Program (Long-Term postdoctoral fellowship LT000733/2009-L), Biotechnology and Biological Sciences Research Council (grant number grant BB/ K009133/1), Isaac Newton Trust (University of Cambridge) and Balfour-Browne Fund (University of Cambridge).

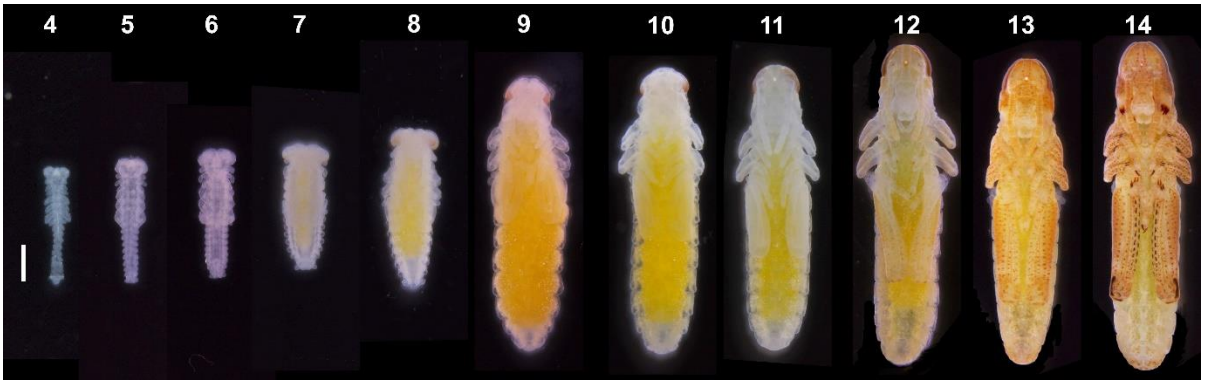
3.10. Author's contributions

BK initiated the study, designed research, carried out all experimental work, supervised the bioinformatics analysis, interpreted the data and wrote the paper; EB performed majority of the bioinformatics analysis and edited the draft; AC carried out the initial steps in the selections of transcripts for the reference transcriptome and did a preliminary expression analysis. All authors read and approved the manuscript.

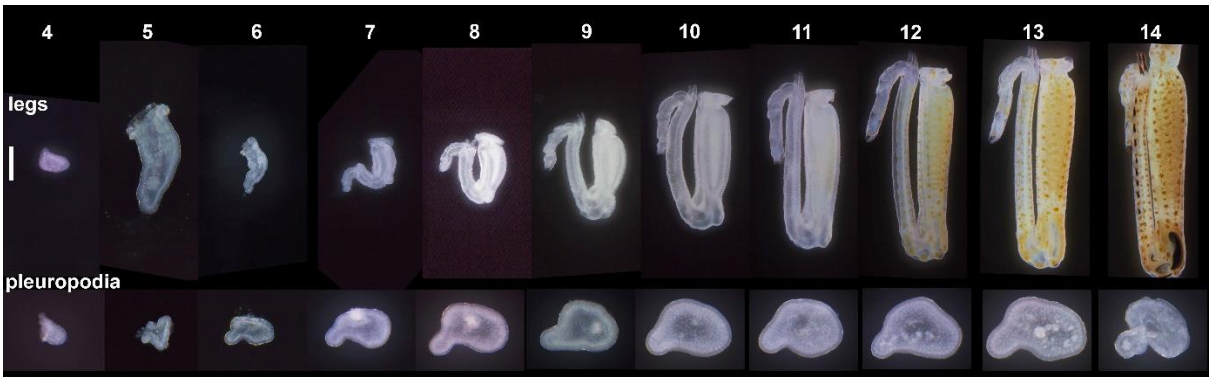
3.11. Acknowledgements

Majority of the work was carried out in the lab of Michael Akam (University of Cambridge) and the data analysis was finished in the lab of Gregor Bucher (University of Göttingen); BK thanks to both for hosting and financial support. Electron microscopy was done at the Cambridge Advanced Imaging Centre (University of Cambridge). Immunolabeling was done in the lab and with help of Andrew Gillis. Stereomicroscopic pictures were taken in the lab of Paul Brakefield. We also thank for help and advice to Ken Siggens, Jenny Barna, Jeremy Skepper and lab, Steven Van Belleghem, Barry Denholm, Jan Sobotnik, and Gareth Griffiths, for scripts to Erik Clark and Simon Martin. We thank to Michael Akam, Siegfried Roth, Stuart Reynolds, Nico Posnien and Maurijn van der Zee for comments on the manuscript.

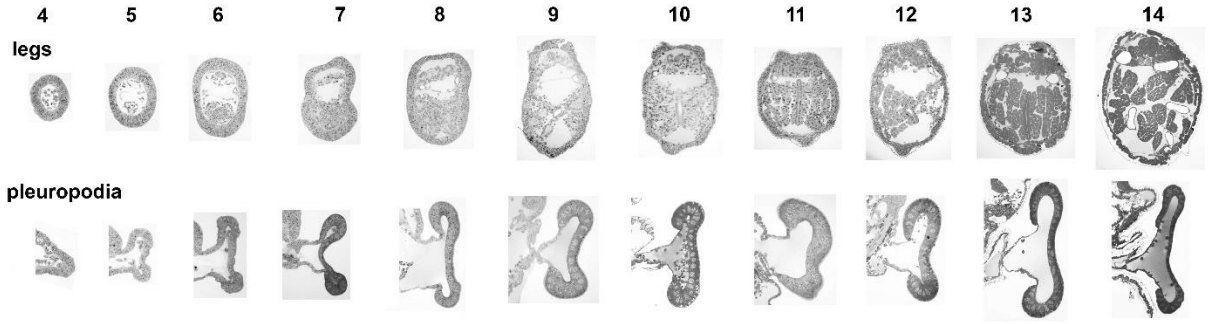
3.12. Supplementary Figures



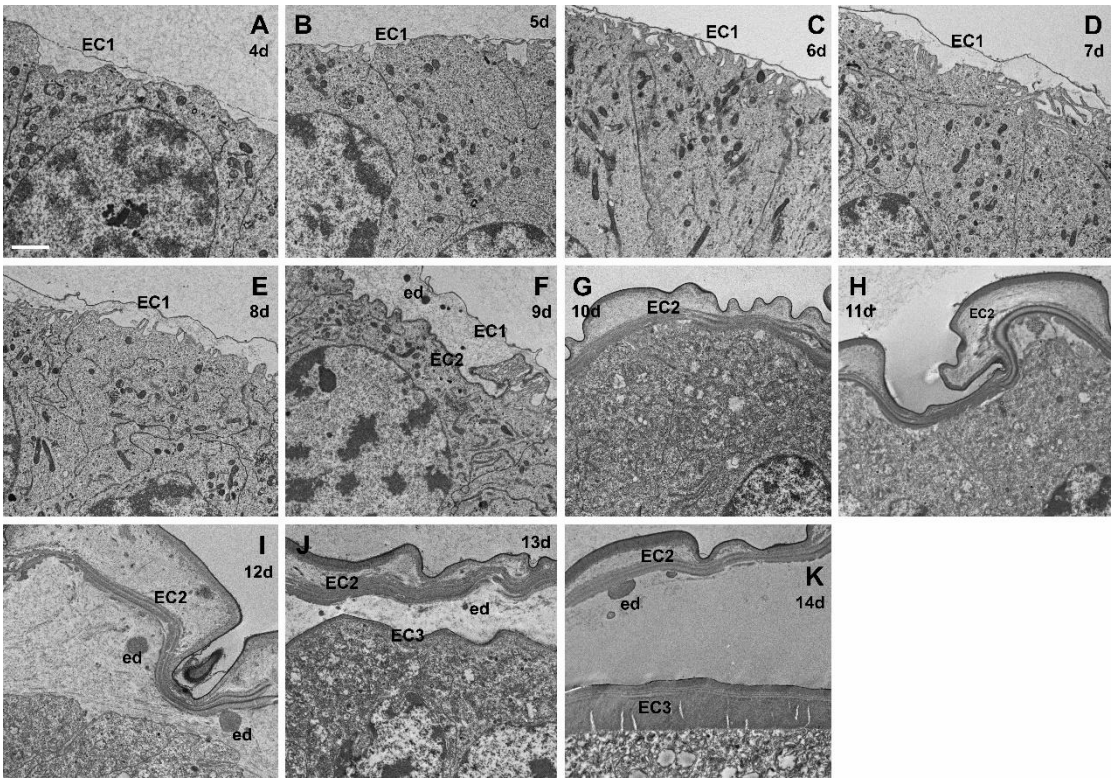
Supplementary Figure 1. *Schistocerca* embryonic stages used in this study. Images of live embryos dissected out of the eggs; imaged under a stereomicroscope. Eggs and embryos of *Schistocerca* typically slightly vary in size. Numbers indicate age in days. Scale bar: 1 mm. Background in photos was cleaned (see Materials and Methods).



Supplementary Figure 2. External features of developing hind legs and pleuropodia. Compare the sizes of the appendages; imaged under a stereomicroscope. Numbers indicate age in days. Scale bar: 0.2 mm for all pleuropodia and for legs at days 4 and 5; 0.5 mm for legs at days 6-14.



Supplementary Figure 3. Figure S3. Cross-sections through developing hind legs and pleuropodia. Toluidine blue stained semi-thin sections of appendages embedded in epoxy resin. Numbers indicate age in days.



Supplementary Figure 4. Ultrastructure of epidermal cells in developing hind legs. TEM micrographs. Compare with pleuropodia in Figure 3. Note the three different cuticles and appearance of ecdysial droplets (ed) during embryonic moulting. EC1, EC2, EC3, the first, the second and the third embryonic cuticle, respectively (EC3 becomes the cuticle of the first instar larva). Scale bar: 2 µm.

(A)

Sg-nag1

MSVISTTVLVFALYGIFSCFATQAEEERPVWTWECRESRCEKVAAGEGEAQSLGACRLSCDPWATLWPRPRGGLQRTPGRLLALNPYSVSVEAAGRDLQP
GVRQLLQEAGRIFHRKVERKARTGAKLRSAGERRSLFVTLTVSDGQTRSFHTDTSEAYSLSISEVTAGRVNAAVTADTFFGARHALETLYQLIVYDDINKQLLL
LSEINLSDSPAFPHRAIALDTARSYFSVASIKRTIDAMAANKLNTFHWHITDSHSFPFVSETFPKLSQYGAYSPEKVYTPDEIKSVVEYARVRGVRIIPEFDAPAH
VGEGWQWVGDNATVCFKADPWSQYCVEPPCGQLNPTSEKMYQVLAGIYKDMLNVFDSDVFHMGGDEVNMNCWNTSEVITDWMDANGIPRTEEGL
HELWDRFQSRAYSLLAEANGKKELPVILWTSTLTDVAHVDKYLDNKRYIIQIWTRGTDLVIPELIRKGFRVIFSNYDALYFDCGFGAWIGSGNNWCSPYIGW
QKVYDNNVWDLLSAFGIDVGEGSEARKLVLGSEAALWSEQADEFALDGRLWPRAAALAERLWTDPVEGWMSAEHRFLIQRQRLVDEGIAADTIEPEWCL
QNQGHCYA*

Sg-nag2

MAPAPPAPHLLALTLLLTLLPSPPVVWANSPRWQWTCDSGLCVRSEAPPEPRLDAELEETVVQRSVHRLRPPWPSHELCRLTCGPYGALWPRPTGHTLIA
DALVPFNPATARFDLSAVAGEQGRELVDAASRRWVRDLQHALAASGGHGGGGEVAGAAAGAGTDVLVTVLTRDSPQALSWETDETYTLDVASSGHEVR
VTVSAQTVWGALHGLTSLRQLVGCCSEDGAALMVAEARIVDGPVYAHRGLLLDTARNFLPVETMMATMDAMAASKLNVLHWHATDSQSFPLLLPRVP
QLARWGAFSARETYSSQQVSALLGYAHARGIRLLLELDAPAHSGQGWQWGEAEGLGALALCVGQQPWRRLCIQPPCGQLNPANPRLVGVLADVYRDVV
DLWPPGQPLHMGGDEVSYSCWNSSAEVLEYMSKRRWDRSQDGFLRLWAEFQQAALEALDAARGSSDVPAILWSSHLTRPGNIERFLNSSRYVIETWVE
GGDPLPQQLLALGYRLVVATKDAWYLDHGFWGSTRYHDWKAVYSNRLPGSMAQGVLGGEVASWGELVDDQSLDARLWPRAAALAERLWSNPGASAR
EAEPRLHAHRARLVAAGVRPEALAPRYCVLNEGACQ*

Sg-fdl

MSRQRLLWRLLGAALALTVAGLAAPPLFRLLVSPHSAANSVAGRRVYSSDPGPWTWSCESGRCVRALWQGGTQVSLDTCQWTCAGWEAPLWPRPTGA
LRLANSTAALPEDLDVRLRLSGPQHEDTRGLLAAATERLARHLQLVRPAWAGRVACDAARGATVARLTVFVKLDADGSRPTGQLTLDTDESYRLQVRRESQ
DLQAEIDARSFFGARHALETLSQLAWWDPVSGCVHILDSAIVKDAPKFRHRGLMVDTARNFIPLEALQRTVDAMASNKLNTLHWHLTDSTSFPYLSRALPT
MARYGAYSPEQVYSMEDVSRLAEFARERGVRLVVELDVPAHAAAGWPTEQVSCSEQRGSAANAPLVQQQQHRQNEDNGLQYRQEERRERRAQHGGE
QQPAWWELCGQPPCGQLPPADEAAFGTLRTLYQELRQASGASDVAHLGGDEVSAECWGGVRGERLWSLWGGFMRRAHRELVAASQGNPPTAVLVW
SSELTAPHNLRRYFDPSTHVVQVWGGSKWNETLPVLLAGFRAVVSHVDAWYLDCGWGDFRSGGPGPCGPVATWQTVYSHRPWAAFPPGARSRLLGGE
ACLWSEKVDDQTLDVRLWPRAAALAERLWSDPPAGVHPDLPPPGSPQRDEPTLRRAYQRLSHHRERLVARGVRAEAMWPRYCHLNPGACF*

Sg-hex

MGKKVEVVLCACVCVGLLLTVTAAEPLPRYITEPGPTVKATQGAVWPKPQNEQRFGGSVLIVPGNFTFQVEGPECDILSEAVSRYEAILKEEAAIKGPRNASE
ASTQLSALLVRLDGECGDRPVFGMDESYELRINSPDLPGAMLLTSASVWGILRGLETFSQVATRVKTADALILDNLAIADIPRFSHRGLLLDTSRHFIPVSYIKK
TLDAMAYNKMNVFHWHIVDDQSFPYQSAAFPLLSEKGSYDPERFVYSPADVAEVIEYARVRGIRVVPEFDTPGHTRSWGEAYPDLLTPCYNATGSPDGTY
GPIDPTKNFTYEFLQTLFEEIVNVFPDEYFHLGGDEVGFECWESNQDILDFMSEHNITESKDLESYYIQKIVDIASNLNSKSIVWQEVFDNEVRLSADTVVHIW
TGDRNEELDSVTAAGHYTLLSQCYYLDRFRYFGGDWHKFYNCEPLDFSADNVYQYDLVIGGEAAMWSEFVDESNVESRVWPRASAVAERLWSPMNVTD
IDEAATRIEEHYCRLRRRGINAQPPNGPGYCV*

(B)

Sg-cht5-1

MRTSAAWFLAVAGLCVVFCPPLVSGNVGDRGRVVCYFSNWAIYRPGIGRYGIDDVPASMCTHLVYSFIGVSNVTWGVLVIDPENDVENHGFANFTALKSK
YPGLKTQLAIGGWAEGGRKYSAMAAVPARRRSLIASVVEYMKRYGFDGFDLDWEYPGAADRGGSFSDKNHFKCFVQELREAFDAEGQGWEITMAVPLA
KFRLQEGYHVPELCELVDAIHVMSYDLRGNWAGFADTHSPLYKRPHDQWAYEKLNVHDGLKLWQDMGCPAHKLVVGVPFYGRSFTLSAGNKDYKLGTY
INKEAGGGKPGNYTQAKGFLAYYEICLEIQEVGGWTEKWDEAGKVPYAYKGTQWVGFENPKSVQIKMDFIKAKGYGGAMTWAIDMDDFRGVCGPKDA
LISVMYNNMKDYIVPDIQYSTTKRPDWDRPPPCDGKKPGAAPASTTTRRPTAAPTQSTTRRPAPTTTAAPSSSSSTTTTRRTTTASRPSTQPPPPPAAPDDN
ELPPAAIDCSDGDFVPHHDCSKYYRCVYGKPVEFSCYEGTVWNPQLRVCDRPNDVHRTDCSMAKLHS*

Sg-cht5-2

MRAATQVGLLLAVALALAAASDEDTTPLDSSTGSPTNSVDEESSSSENAAVLSGGQRRGRVTCYFESWAVYRKRLRYGIEDIPGDMCTHIIYSFVGLNNVT
WELQVLDEKLDVQDGGFENFTALRQEFPGVRLQVALGGWAEGGHNYSAMVGDPARRASLVRSAVAFLHRYGFDGFDVDWEYPGNAPRGGVPEDKDD
FLCFMQELRVAFDAEGLGWELTMAVPLTEDKLRDGFHVPQLCSIVDAVHVMAYDLRGEWDHFADVHSPLYRRPHDTGAYAKINTHDGLLLWEQLGGSS
WGCHSTATPTNCVPTSPTTLPVLASFRAPEMTSAE*

(based on alignment with homologous sequences this transcript might be misassembled and the amino acid sequenced prematurely terminated by introduction of a stop codon)

Sg-cht10-1

MWRPVALSLWLLLATSRGLHVPPADEPSFVRDAVEAPPGQSLALRRSATASRPRLPAFGTRQLPLRQAVESPPMAARLRSSERLPLRDAVEHVPYEALPGA
PTASEAFSLWRGFGDWLPENLPSTRQFNHSFAWWHDAIIAKLSLGGPRTKPPSLQAPSTHTSGIRQFKVVCFVEGWAGYRRDPMRFTTADIDPFACTHIIY
AFAVMDPHDLHIKPQDEQYDIIQGGYRSIVGLKRQNPQLKVMISVGGWPEERRKFAEMTASASTRREFIRSVLHFIDEYGFDGIDLDWEYPGAADMGGSA
REKEHFSLLVEELAEAFAPRGSVLSASVSPSRFRVEDGYDVPRLARRLDFLNLMAFDLLTEQDAAADHHAPLTQRKHDYGLAVFYNVDYAVRYWLRKGARR
DQLVVGIPFHGHSFTLQDEAKNSPGAPVKGLGKEGPYTQEKGFLAYFEILQLLEEGHWMKATDDVGSPYMVKGNQWIGYEDQRSIATKVMYIKKNLLGG
AMVWALDLDDFEGAYGQKWPLLSVVKKGLLETTPQSDQQQASQEPTHVTPPIAGVPVSVDSSQYNCSGRGYVRDSASCQIYHRCEWGMKHTYICPEGL

HYDSRTQLCDWPQIANCPMDNSSQRIEQENQSEVACNEEGLMEDPKDCNRYYMCHKGVAQHYSCMLGQYFNVQKGICEYGSCMPKAPQDNIPSSQTR
NLVGEDHYKVVCYYASWAWYRKEGGKFVPEHIDPTLCTHIVYAYASLDPNTLTMKYFDERADKENNFYERLTELPKKSGHHQQHASDVTVMIGLGGWTD
SAGDKYSRLVSEGSARRRFVSKAVEFLHRHQFGGLHLDWDYPRCWQSNCGRGPTSDKPNFTKLVQELRQAFKKQNPPLALAISISGYHEVIDEAYDLAELG
RNTDFMSVMTYDYHGSWEKSTGHVSPLYHRNGDIFPMYNTNDTMEYLVNKGAPRDKLLVGIPFYGQSYTLENPSNHDIGAPATGPGLAGEFTMQPGML
AYYEICDRVRNNFWKIGRDRFGATGPFAYAGNQWVSFEDTKSVKEKAKYIKNMGYGGAMTFTLDLDDFENRCCRGAFPLLRSINRVFGRIPDSAEPSGDD
CTRPPPPVTPPPPTYTTGVDSGDHRPTTPISTTHQHPTSPKPSTTEYPWW

Sg-cht10-2

PSTTTSTTTSTTTTTTTTTTSTTTTPRPTTRPTTMSTTEYPWWTPSTTSTTRKPPTTRPTTTSTTEEYPWWTPPSTTKKPTTSSTTEQPWWTPSTTSTTSTAAP
TTTMTTTEKPWWSTTPQKPLPPDSGPCEAGVYYPDPTNCNAYYRCVLGELRKEFCAGGLHWNPDKKVCDWPSESKCDTKEPSETTVGSTTSSTTENPWW
TPSKPSETQATTTTTEVPWWSTTRPPRPPTTEGNSEWVTTSRPTTTQQPSEEVSECMNGQYYPVAGSCKSFYICVNGRLIKQTCAPGLVWNQDQTMCD
WGFNVKCADDSEREAVHKAQPDDPCNQGALNPYPGDCTRYLYCQWGRYHEADCAAGLHWNEMEKICDWPENAKCTDMESGSEAPAASSQKPVTEM
STSWTTAAPTTKPPWTWATTTTVKPVTTTSTRAPPAQGPPISGYFKVVCYFTNWAWYRRGLGKYVPEDIDANLCTHIVYGFAVLDYENLIIKAHDSWADFD
NKFYERVVAYKKKGLKVSLAIGGWNDSAGDKYSRLVNSPSARRRFIKHVLEFLEKYGFDGLDLDWEYPVCWQVDCAKGPASDKSSFAALVKELRQAFEPKG
LLLSSAVSPSKTVIDAGYDVKTLAENLDWIAVMTYDFHGQWDKKTGHVAPLYFHPDDDFYFFNANFSINYWISEGAPRRKIVMGMPLYGQSFQLEKASTN
GLNARSTGPGQAGEFTRAAGFLAYYEICDRIKNKGWTVVQDPERRMGPYAFKGNQWVSFDDVAMIQQKSEYIRKMGLGGGMIWALDLDDFRNRCGG
GTHPLLNTIRTVLAAPPGGDGATEMPPSWSTPGGGQPTMSTEEWMSSTSISSTEITDSGHHSTQDSGGEVTSVSPAITTTNRPAHPGTSSSPPPPPSQGEEF
KVVCYFTNWAWYRQGVGKYLPNEIDPDLCTHIVYGFAVLNGDRLTIKPHDTWADYDNKFYEKVTEYKKKGIKVLVAIGGWNDSAGDKYSRLVNSPGARRR
FIEDVIDFIEQNNFDGLDLDWEYPKCWQVDCKKGPDSDKEAFAAFVRELRAAFNPKGLLLTAAVSPSKAVVDAGYDVPTLSQNLDWIAVMTYDFHGQWD
KITGHVAPMYTHPEDVDVTFNANFSIHYWIQKGASPKKIVMGMPMYGQSFSLADNSDHGLNAPTYGGGEAGESTRARGFLSYYEICTNIQKKGWRVVKD
PEGRMGPYAYLRDQWVSFDDTSMIRYKSNFIRRMGLGGGMIWALDLDDFRNVCSCEKYPLLKTINRVLRGYPGPGPNCDIEATEKPGSEETDNRIHPTIPP
TSSTNNWNVISGGGGIVPKDPTCGNRLFAPHDKDCNKYYLCQYGDFMEQSCPQGLYWNKDHCDWPSNTDCSKEDSSVINPAPIASTQEPEMSSTTENIH
MSESTVTTSIRPSEPGTSTVMTPSGDYMVVCYFTNWAWYRQGLGKYLPSDIDTSLCTHIAYGFAVLDGNSLTIKPHDSWADLDNEFYTKVSGLKKKGIKVLL
AIGGWNDSLGDKYSRLANNPSARRKFVEHVVKFIEKYGFEGLDLDWEYPKCWQVDCNAGPDSDKQGFADLVKELSMAFKPRGLLLSSAVSPSKVVIDSGY
DVPVLSQYFDYISVMTYDFHGHWDKQTGHVAPLYYYPGDTYDYFNANFTMHYWIEKGADRKKLIMGMPMYGQSFSLADAKNHGLNAKSYGPGEAGEF
TRAGGFMAYYEICYNVKSKGWTTVRDPEGRIGPYAYRGNQWVSYDDVSDIRRKTQFIKELGLGGGMIWALDLDDFRNRCGCGTYPLLRTINSELRGLTAN
THDCT*

Sg-cht7-1

MIAPRCVWRAALWCVVIILLADLVYSASSTGRRRLRRPGGSSSSSTTSSSSSTSTKVRTRDQETSASVNRFRVRNRLTPPGANRKSGSGSAVAAASDKSGGY
KVVCYYTNWSQYRTAHGKFLPEDITPDLCTHIIYAFGWLKKGKLTSFEGNDETKDGKVGLYERVMALKKANPKLKVLLALGGWSFGTQKFKAMSETRYTRQ
TFIYSAIPYLRKHDFDGLDMDWEYPKGTDDKKNFVLLLKELREAFEAEAQEVKQSRLLLSAAVPVGPDNVRGGYDVPAVASYLDFINLMAYDFHGKWERET
GHNAPLYAPSSDSEWRKQLSVDHAATMWVKLGAPKEKLVIGMPTYGRTFTLSNPSNFKVNAPASGGGKAGDFTKEGGFLAYYEVCDMLKKGATYIWDD
EMKVPYAVMGDQWVGFDDERSIRHKMKWLKEGGYGGAMVWTVDMDDFTGTVCGGGVKYPLIGAIREELRGVSRGPNAKDVDWSKVARTVSLEATT
KPAPIKIDVSEVLNRVRKPTKQAPADLSNEVIDLNSRPAQVFCYMTSWSGKRPGAGKFSPEDVDPSLCTHVVFAFATLKDHKLAPANDKDDGLYERVIALRE
KNPQLKVLLAIGGWAFGSTPFKELTSNVFRMNQFVYDAIELLRDFKFDGLDVDWEYPRGADDRAAYVSLLKELRMAFEGEAKTAEQPRLLLSAAVPASFEAI
AAGYDVPEISKYLDFINVMTYDFHGQWERQVGHNSPLYPLESATSYQKKLTVDFSAREWVKQGAPKEKLLIGMPTYGRSFTLVDTSKFDIGAPASGGGAAG
RYTAEAGFMAYYEVCDFLHHDNTTLVWDNEQQVPFAYRGDQWVGFDDERSLKTKMGWLKELGFGGIMVWSVDMDDFRGQCGAGKYPLLTSMRQEL
RDYRVQLEYDGPYESRGPLGAYTTKDPTSVSCEEEDGHISYHPDKADCTMYYMCEGERKHHMPCPSNLVFNPNENVCDWPENVEGCMHHTQAPPAAR
RR*

Sg-cht7-2

MTWPPPPLLLSLLVLLATSASARFVSTHDVTPCAVEALAPSDKALLCYYEGRLSVYQLDPCLCTHIVFKDAAVVSDNFGLKIVSDVSGASLLRARSPSLRTVLGL
RLSGAVARAALASPSRRLALARDAARRLYAHHLDGIELSVDDDEAASAAAADAAPAATARQGLVALLKALRTALDSHGREKRDYLVSEQVFDDFTTQEYEPT
WSDGSSRKSRRRATTTTTTTSTTESPEETAARYLELERDAQNAQLLLSLPTKPETIAKRYDVKNITRYVDYVVLRTQAMTDDSERGLVYHPSRLMGLDDMLN
ADAVVDLVTSLGASPAQLVITLPGQATAFELRREDRTEPRSPASGAPRTISQPELCRALSRGNWTLERDEDQTAPYAYSGRRWIAFDDALSASIKGKYAVVR
GLAGTAVDAADALDWQGTCGAPASQLRALHSALAQLRRSSRGALLHGLE

Sg-cht7-3

DKGMPKNKIIVGIPTYGHSFRLINAENHGWSAPASGYGKIGSKGFVSYPEVCQFLHSTGSKYIFDKNFEVPYAYQGLEWISYDDECSVMYKAKYIASSSYGGA
MVFSLNVDDHQGVCAGTTFLLTTQIRNILGVSWQ*

Sg-cht2

MQQLAPLAFVLAFLAAAFAASPLGHNKAVVCYVSSWAVYRPGNGVFTVSDINPNICSHLVYAFAGLNATDNTIITLDKYNDLEEDYGKGNYKKITGLKNQYP
HLKVSIAIGGWNEGSANYSHMASTPTTRQQFIRSVVNFLRKYNFDGLDLDWEYPTQRGGVPSDRENFVALVRELRQEFDKNGWLLTAALGASTAVIEKAY
DVPMLGKYLDYMHIMCYDYHGTWDKMTGANAPLYGSSPSDTLSVDNSIRYYLKLGAPAKKLLMGVPLYGRTFMSDANANMGGLGAPAEEKSFQGPYTK
EDGYMGYNEICLELKTNSSMWTIMWDDKSSTPYAVSTNKVIVYDNAKSLTEKVNLAMKLELGGIMVWPLDTDDFRGECSEGIYPLMHTINKAIVQSSQQK
SDSSGMKVPDSTAAASCGCASLIFLSFLYLFQL*

Sg-cht6-1

VCYYTNWSVYRPGTAKFTPQNINPYLCTHLIYAFGGLSRENGLRPFDKYQDIEQGGYAKFTGLKTYNKDLKTMLAIGGWNEGSTRFSPLVADAERRKEFVKN
VLRFLRQNHFDGLDLDWEYPAFRDGGKSRDRDNYALLVKELREEFDRESEKTGRPRLLLTMAVPAGIEYIDKGFDIASMNKHLDFMNILSYDYHSAFEPAVN
HHSPLYSMEEDDEYNFDAQLTIDHTVNHYMKSGADRNKLVLGIPTYGRSYTLFNPLATELGSPADGPGEQGDSTREKGYLAYYEICENLQSDDWKVVQPN

PSAMGPYAYKGNQWVSYDDMDIIKKKAQYVNDNGLGGIMFWAIDNDDFRGKCHGRPYPLIEAGKEAMLKGVKRSNNEIETTPVQNNRQSSRKRNRNR
SKGNARGRTRTTASTSTVVTTTTTTTTTTAAPLITPSYTTPEPPTTPDPGSDFKCKDEGFFPHPRDCKKYFWCLDSGPSNLGIVAHQFTCPSGLFFNKAADSC
DYARNVVCNKKSKSQGGSSSTLPPIKAATSSTTRFSTSPSTKLTTKLTTTTTTEPPPVLDDDDDDD

Sg-cht6-2

MNIRVKQPVIIGNCYRGQPNRLWEVFILKWFLVAVACLIAAGAVTVYLAHYFMKTRYTSTNVTGVTGQHSDLNTYKGQLQDMGDGYSLFKQEDMTQICK
TDELTGSQQMRKQSTKLVCYYTFPGPGGLVPDKIDPFLCTHINIAAVGINNSKLEPLCEERKEVIKSLVGLKTRNKNLKVILSVIGMPGGFGDMVSKSSSRRM
FIKDL

Sg-cht8-1

MSHFWLRLAVILGVSLSICGAEDKKVVCYHGSWSAYRNGNGRFEIEYIQPELCTHLIYTFVGITSAGEVRILDEWLDLPSGKNAYNRFNALKSSNTKTLVAIGG
WNEGSATYSAVMNDASLRAKFVQNVVNFVKTYGFDGFDLDWEYPANRGGSPGDLTAFVELIKELRTEFDKYGYLLTAAVGVGRYLIGTAYDVPQISKYLDF
INLMTYDLHGSWDGKTGQNAPLYASSADKTEAERQLNVDSSVRYWIQNGADPSKLVLGMGTYGRTFTLSSAANTGVGAPATAPGTNGPYTMESGMMG
YNEICEKINAGGWTVVWDEEQKVPYAVNGNQWIGYDNEESIRLKSQYVLDMGLAGGMIWSLETDDFKGLCGSKTYPLLSTINEVLRGITSTNSGSSSSSSSS
SSSSSSSSSSSSSSSSSSSNTAASASSSSGVCSSAGYVRDPSDCGVFYLCTASGSGYTASKFTCPGDLVFDESSSACNYKSLVAC*

Sg-cht8-2

MSPFLSGLLLLLGVLNICGADEKKVVCYHGSWSAYRNGNGRFEIEYIRPELCTHMIYSFVGITSAGEVRILDEWLDLASGKNAYNRFNKLKSSNTKTLVAIGG
WNEGSATYSAVMNNAALRQKFVQNVVNFVKTYGFDGFDLDWEYPANRGGSPGDLRAYVELLKELRAEFDKHGFILSAAVGVGRYLIGSAYDVPQLSKYLD
FINL

Sg-cht8-3

ESGMMGYNEICEKIKAGGWKVTWDDEQKVPYAVSGNQWVGYDNEESIKLKSQYVLDMGLGGGMIWSLETDDFKGVCGAGTFPLLSAINQVLRGAAAT
SSAGSSTSGSSSGSSGSSSSSSASSGGSSSGTSSGSSTTSSGASADSSGSSASSGSSSSGSSATSVSSGSNSSGVCNSAGYARDPSDCGV

Sg-idgf-1

MAELPLLLLLLAAAATCWTSAAALGATRVVCYLDGGALRRPEPHRMLVSEIEPSLTYCTHLIYGYATIDTDSYKAVPRHEGEGTNYTSVVALKRRFPALNVLLS
IGGGSADSGQREKYLHLLESDEHRRTFVKSAKDLLKQYHFDGIDIAWEFPMNKEKKERSTLGSFWHGFKKVIGLAHSHKDEKADEHRREFSSLIQELKTSLKT
ENALLTLSVIPYINHTLYYDCSALSPHVDHLHLLAYDYHTPQRTPNTADYPAPLYVAGKRDPDLTADGNVRWFLERGFPSRKIILGIPTFARTWKLDDDSRVS
GVPPIEADGAGDTDNIANTAGIMAFQTVCMLLPNAGNAGYKTTLSRVTDPTDRLGSYGFRLPSGEVTGLWVGYEDPDVAQYKAAYAKIKSLGGIAFSDLSL
DDYHGICTGDKYPIVRAGTLKLRYK*

Sg-idgf-2

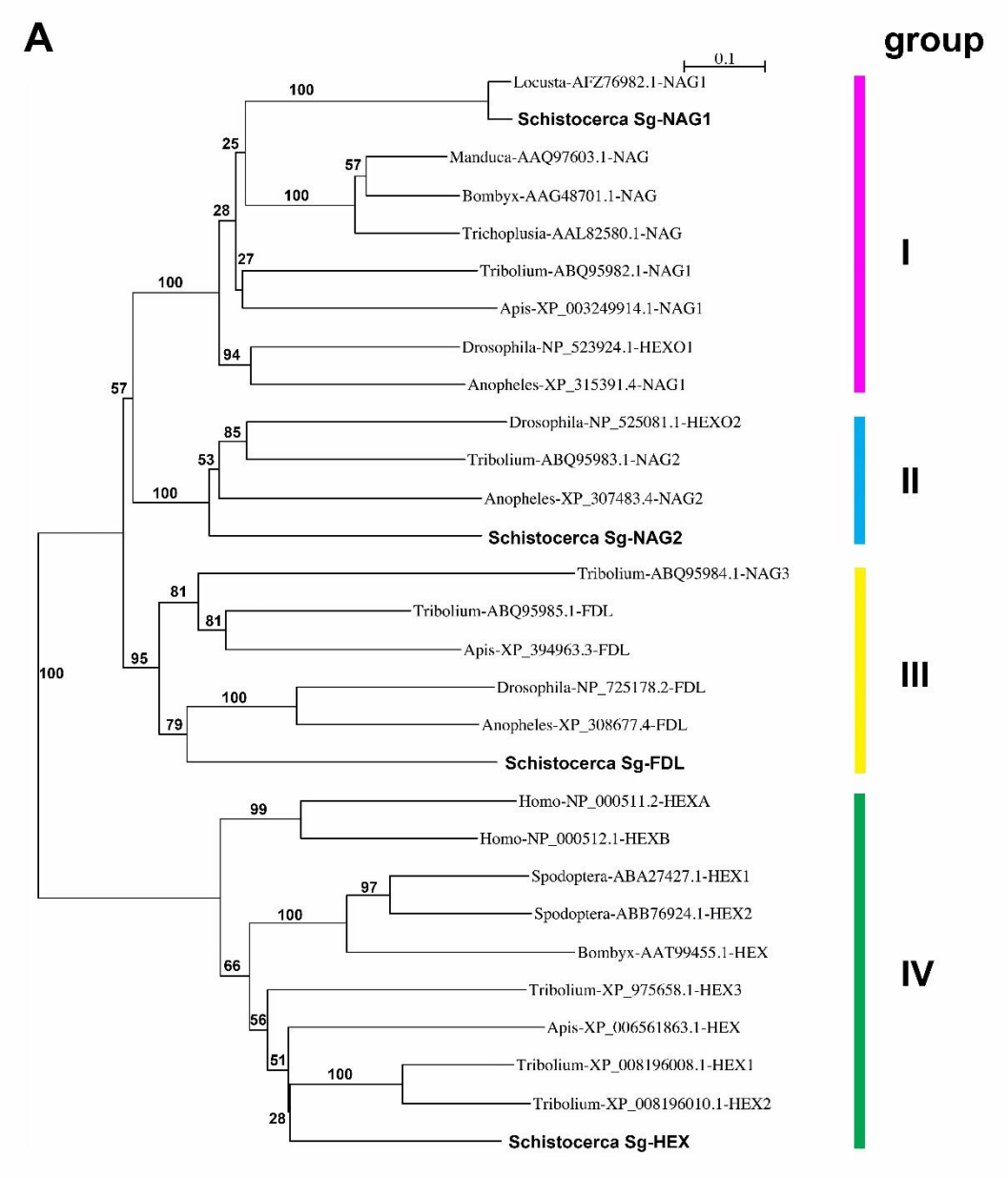
MQSFARLLLLSACCWSAALAATTKVVCYFNTSALKRPESSRMLLSQIEPSFSYCTHLVVGYATINTETYKAVPPSEDEHTTYTNIVALKRRFPSLKILLSIGGGAA
DTDTREKYFELLESDEHRTTFVSSAKSLLKQHGFDGIDIAWEFPKNKAKKDRGTFGSIWHGIKKAVGAAHSHTDEKADEHKSQFSALIRELRTSLRNENALLTL
SVIPYINQSLYYDPTALNQQIDELHVLAFDYRNPERDSQGGRLPCAALPSRAEGLRPLGRREHPLVPRELIPS*

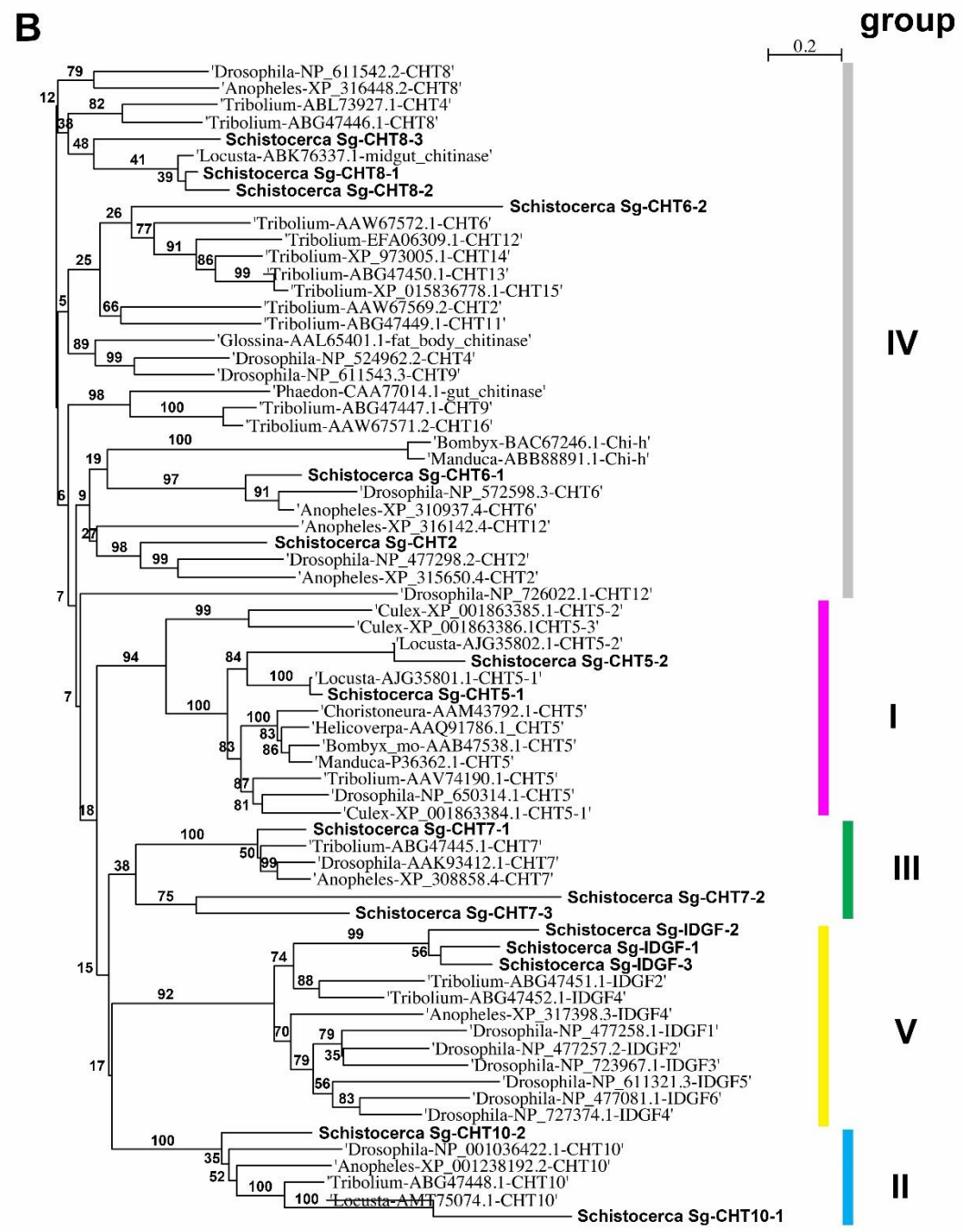
Sg-idgf-3

TDSYKAVPRYQDDTTKYTSLVALKERFPSLKVLLSIGGGGADADQRKKYLELLESDEHRRTFVDSVKELLQQNRFDGIDIAWEFPFSKEKKDRGNVWHGVKKVLGY
AHSHRDENPDEHRRQFSALIRELKSSLKTQNALLTLSVIPYINHSLYYDCASLSPEIDQLHLLAYDYHSPTRTPKKADYPAPLYRAGERSADLTVDGNVRWFLEKGFPS
RKIILGIPTFARTWKLTKDSRITGVPPIDADGPGVAGSIANISGLLAYQTVCTLLPNDANAAYRTTLRRVTDPTDRLGSYGFRLPTREVSGLWVGYENQHSAEYKAAY
ARKKSLGGIAFSDLSLDDYNGVCTGEKFPIVRAGTLKLLSTSV*

Supplementary Figure 5 Amino acid sequences and conserved domains of *Schistocerca* chitin degrading enzymes. A. NAGs, **B.** CHTs. Signal peptide and transmembrane region identified by Phobius (http://phobius.binf.ku.dk/index.html) and conserved domains identified by SMART (http://smart.embl-heidelberg.de/) are underlined and coloured. In **A.** and **B.** signal peptide: magenta, transmembrane region: dark blue. In **A.** Glycohydro 20b2 domain (N-terminal domain of the eukaryotic beta-hexosaminidases): light green, Glyco hydro 20 domain (glycoside hydrolase family 20 catalytic domain): grey. In **B.** Glyco 18 domain (catalytic domain): light blue, Chitin-binding domain type 2 (ChBD2): green; catalytically critical consensus sequence in the Glyco 18 domain, FDG(L/F)DLDWE(Y/F)P, is highlighted in yellow and amino acid changes from the consensus are coloured in orange.

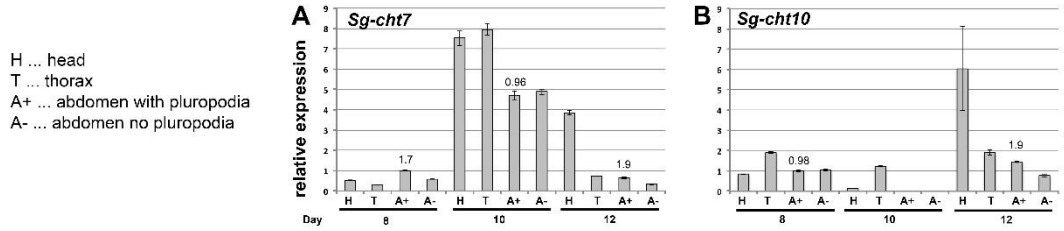
FIGURE S6


A
group
0.1
Locusta-AFZ76982.1-NAG1
Schistocerca Sg-NAG1
Manduca-AAQ97603.1-NAG
Bombyx-AAG48701.1-NAG
Trichoplusia-AAL82580.1-NAG
Tribolium-ABQ95982.1-NAG1
Apis-XP_003249914.1-NAG1
Drosophila-NP_523924.1-HEXO1
Anopheles-XP_315391.4-NAG1
Drosophila-NP_525081.1-HEXO2
Tribolium-ABQ95983.1-NAG2
Anopheles-XP_307483.4-NAG2
Schistocerca Sg-NAG2
Tribolium-ABQ95984.1-NAG3
Tribolium-ABQ95985.1-FDL
Apis-XP_394963.3-FDL
Drosophila-NP_725178.2-FDL
Anopheles-XP_308677.4-FDL
Schistocerca Sg-FDL
Homo-NP_000511.2-HEXA
Homo-NP_000512.1-HEXB
Spodoptera-ABA27427.1-HEX1
Spodoptera-ABB76924.1-HEX2
Bombyx-AAT99455.1-HEX
Tribolium-XP_975658.1-HEX3
Apis-XP_006561863.1-HEX
Tribolium-XP_008196008.1-HEX1
Tribolium-XP_008196010.1-HEX2
Schistocerca Sg-HEX
I
II
III
IV



Supplementary Figure 6. Phylogenetic trees of chitin degrading enzymes in Schistocerca and other insects. A. NAGs, **B.** CHTs. Schistocerca sequences are in bold. Amino acid sequences were extracted from NCBI GenBank. The numbers above the branches are bootstrap support. The marker shows a branch length. Both trees are unrooted. The tree in **A.** was prepared using the SeaView software (version 4.6.1; (Gouy et al., 2010); http://doua.prabi.fr/software/seaview): alignment with default parameters, tree using the Neighbor Joining method, Poisson distribution, 5000 bootstrap replicates. The tree in **B.** was prepared using the CLC Sequence Viewer (version 7.8.1; https://www.qiagenbioinformatics.com/products/clc-sequence-viewer/): alignment with default parameters except gap open cost 3.0 and gap extension cost 3.0, tree using Neighbor Joining method, Kimura model, 1000 bootstrap replicates.

FIGURE S7



Supplementary Figure 7. **Real-time RT-PCR expression analysis of Sg-cht7 and Sg-cht10-1 on cDNA from parts of Schistocerca embryos.** cDNA was prepared from mRNAs isolated from parts of embryos at the age of 8, 10 and 12 days: H, head; T, thorax; A+, abdomen with pleuropodia; A-, abdomen without pleuropodia. Analysis of 3-4 technical replicates is shown. Expression in A+8 (abdomen with pleuropodia when they first become differentiated) was set as 1. Numbers above A+ expression is fold change from A- of the same age.

3.13. Supplementary Tables

Supplementary Table 1. Embryonic transcriptome of *Schistocerca*: numbers of sequenced reads and assembled transcripts.

Samples[a]	Reads total	Unique transcripts	Transcripts in reference transcriptome[b]
1-4d embryos	96,907,644	70,529	20,834
5-7d embryos	92,825,202		
8-10d embryos	99,198,014		
11-14d embryos	96,759,706		
8-9d legs	38,919,230	40,143	
8-9d pleuropodia	22,302,378		

[a] in "embryo" samples the mRNA was isolated from whole eggs collected at each day, then in indicated age groups pooled together for sequencing

[b] see Materials and Methods how transcripts for the reference transcriptome were selected

Supplementary Table 2. RNA-seq expression analysis: numbers of sequenced and mapped reads.

Sample	Reads total	Reads mapped
4d LEG	50,592,896	38.404.015 (75.91%)
4d PLP	47,004,156	35.905.385 (76.39%)
5d LEG	49,391,167	35.559.693 (75.11%)
5d PLP	49,002,608	36.095.324 (73.66%)
6d LEG	50,647,001	37.684.851 (74.41%)
6d PLP	49,111,150	37.490.747 (76.34%)
7d LEG	47,410,277	35.958.856 (75.85%)
7d PLP	47,275,171	35.971.381 (76.09%)
8d LEG	49,998,624	38.119.439 (76.24%)
8d PLP	48,420,404	37.706.738 (77.87%)
8-9d LEG	38,919,230	29.467.879 (75,72%)
8-9d PLP	22,302,378	16.152.357 (72.42%)
10d LEG	49,170,085	37.814.977 (76.91%)
10d PLP	46,901,233	35,403,192 (75.48%)
11d LEG	49,472,441	37.815.815 (76.44%)
11d PLP	48,516,135	36.818.833 (75.89%)
12d LEG	47,068,033	34.117.674 (72.49%)
12d PLP	46,801,370	34.936.272 (74.65%)
13d LEG	46,658,116	33.454.889 (71.70%)
13d PLP	49,776,232	37.167.588 (74.67%)

Supplementary Table 3. Number of differentially expressed genes at selected levels of stringency.

Day		4		5		6		7		8	
RPKM	**Fold change**	**DOWN[a]**	**UP**	**DOWN**	**UP**	**DOWN**	**UP**	**DOWN**	**UP**	**DOWN**	**UP**
>10	**>2**	29	19	77	195	360	589	649	857	944	791
>50	**>2**	5	6	18	63	97	241	181	394	289	403
>100	**>2**	2	3	7	26	31	130	70	238	111	265
8-9		**10**		**11**		**12**		**13**		**10+11+12**	
DOWN	**UP**	**DOWN**	**UP**	**DOWN**	**UP**	**DOWN**	**UP**	**DOWN**	**UP**	**DOWN**	**UP**
890	850	1538	857	1874	842	1358	772	1196	871	781	1535
259	430	427	454	457	411	492	408	523	478	451	484
108	256	215	301	216	292	287	286	350	312	327	277

[a] DOWN: downregulated, UP: upregulated

Supplementary Table 4. Differential expression of genes, whose expression dynamics in the early stages is known.

			Day 4					Day 5				
Gene	**Predicted expression[a]**	**Transcript ID**	**RPKM LEG**	**RPKM PLP**	**prob**	**UP[b]**	**DOWN**	**RPKM LEG**	**RPKM PLP**	**prob**	**UP**	**DOWN**
Ubx	high in pleuropodia, lower in legs	SgreTf0014307	58.98804099	176.4998914	0.752327603	2.992130074		71.70784725	188.2842414	0.706848869	2.625713205	
abd-A	low in pleuropodia, not in legs	SgreTf0002957	0.127286025	17.22135392	0.78513135	135.2965023		0.098924125	9.139851296	0.764717385	92.39254151	
Dll	similar in both	SgreTf0013577	33.38516081	42.26190639	0.3806017	1.265888957		30.76566587	34.08765844	0.296615141	1.107977269	
dac	low to no in pluropodia, high in legs	SgreTf0002755	30.55098317	2.435815478	0.79098899		12.54240456	46.45842693	3.305435244	0.796238011		14.05516173
abd-B	not in either pleuropodia or legs	SgreTc0000012	filtered out - no expression in legs or pleuropodia									

[a] Tear et al., 1990 (abd-A); Kelsh et al., 1993 (abd-B); Bennett et al., 1999 (Ubx); Beermann et al., 2001 (Dll); Prpic et al., 2001 (Dll, dac); Hughes and Kaufman, 2002 (Ubx, abd-A, abd-B); Angelini et al., 2005 (Ubx, abd-A, abd-B); Zhang et al., 2005 (Ubx, abd-A); reference list is in Additional file 1

[b] significant upregulation (UP) or downregulation (DOWN) are highlighted in magenta and blue, respectively; threshold for differential expression: prob > 0.7, RPKM > 10, fold change > 2

Supplementary Table 5. Comparison between differential expression of selected genes obtained by RNA-seq and real-time RT-PCR.

		Real-time RT-PCR[a]		RNA-seq[b]		Details RNA-seq		
Day	Transcript ID	UP	DOWN	UP	DOWN	RPKM leg	RPKM plp	prob
4	SgreTa0007432		1.058	1.064		11.177	11.888	0.261
4	SgreTa0001469	1.303		1.214		96.102	116.656	0.359
4	SgreTa0005616	not detected		6.282		0.002	0.012	0.261
4	SgreTa0013453	1.734		1.543		75.079	115.851	0.521
4	SgreTa0008219	not detected			3.584	0.232	0.065	0.261
4	SgreTa0001661	1.058			1.035	38.974	37.662	0.265
4	SgreTa0014626	1.055			1.022	145.383	142.194	0.263
5	SgreTa0007432	1.099		1.011		11.303	11.430	0.266
5	SgreTa0001469	2.060		1.510		84.745	127.987	0.515
5	SgreTa0015941	210.358		168.642		0.276	46.467	0.797
5	SgreTa0007802	3.726		5.914		0.450	2.659	0.543
5	SgreTa0005616	UP indefinitely		2.473		0.056	0.138	0.266
5	SgreTa0017664	not detected		NA[c]	NA	NA	NA	NA
5	SgreTa0009118		1.823		1.972	119.291	60.507	0.579
5	SgreTa0000088	1.074			1.171	55.737	47.592	0.333
5	SgreTd0002755		10.247		14.055	46.458	3.305	0.796
5	SgreTa0001341		8.662		10.941	22.491	2.056	0.790
5	SgreTf0013577	1.015		1.108		30.766	34.088	0.297
5	SgreTa0005600	UP indefinitely		5.204		0.526	2.739	0.543
5	SgreTa0013453	2.469		2.368		81.519	193.025	0.689
5	SgreTa0008219		1.123	2.095		0.314	0.657	0.266
5	SgreTa0008219		1.120	2.095		0.314	0.657	0.266
5	SgreTf0014307	2.661		2.626		71.708	188.284	0.707
5	SgreTa0001661	1.237		1.238		38.483	47.658	0.360
5	SgreTa0014626	1.427		1.408		142.712	200.902	0.454
5	SgreTa0007477	4.762		5.104		52.030	265.551	0.789
6	SgreTa0007432		1.181		1.258	10.152	8.069	0.369
6	SgreTa0001469	2.475		2.182		85.763	187.142	0.680
6	SgreTa0002409	30.406		33.056		8.717	288.152	0.794
6	SgreTa0015941	12.189		22.566		0.040	0.907	0.289
6	SgreTa0007802	3.463		3.248		0.809	2.627	0.448
6	SgreTa0005616	UP indefinitely		901.151		0.198	178.706	0.971
6	SgreTa0017664	UP indefinitely		2717.495		0.211	572.225	0.999
6	SgreTa0009118		3.750		4.200	117.188	27.902	0.776
6	SgreTa0000088		1.065		1.136	46.963	41.325	0.329
6	SgreTd0002755		320.639		275.795	74.322	0.269	0.794

6	SgreTb0006243	UP indefinitely		2.287		0.442	1.011	0.289
6	SgreTa0017707		8.595		2.935	4.258	1.451	0.528
6	SgreTa0017736	8.122		7.221		0.156	1.129	0.289
6	SgreTa0008528		DOWN indefinitely		1.681	3.778	2.247	0.396
6	SgreTa0005600	UP indefinitely		97.272		1.880	182.830	0.794
6	SgreTa0013453	2.602		2.513		76.870	193.188	0.707
6	SgreTa0013453	2.602		2.513		76.870	193.188	0.707
6	SgreTa0008219		2.208		1.516	0.752	0.496	0.289
6	SgreTa0008219		1.327		1.516	0.752	0.496	0.289
6	SgreTf0014307		1.193		1.082	112.101	103.605	0.308
6	SgreTa0001661	1.488		1.259		40.953	51.580	0.398
6	SgreTa0014626	2.116		1.868		162.842	304.203	0.585
7	SgreTa0007432		1.035		1.217	10.039	8.247	0.347
7	SgreTa0001469	5.404		2.822		82.172	231.897	0.718
7	SgreTa0007802	563.365		300.262		1.148	344.621	0.795
7	SgreTa0005616	UP indefinitely		619.842		0.242	149.935	0.927
7	SgreTa0017664	UP indefinitely		251.604		0.330	83.089	0.794
7	SgreTa0009118		4.507		6.370	98.857	15.520	0.791
7	SgreTa0000088	1.266			1.005	45.617	45.398	0.303
7	SgreTa0014975	4.777		3.028		17.526	53.072	0.752
7	SgreTb0019973	87.226		63.539		18.121	1151.375	0.795
7	SgreTb0006243	1268.530		524.381		0.555	290.855	0.927
7	SgreTa0017707	52.614		25.430		18.890	480.378	0.795
7	SgreTa0017736	54.154		16.270		0.719	11.703	0.769
7	SgreTa0007897	391.606		160.119		0.530	84.880	0.794
7	SgreTa0008528	229.010		147.348		3.858	568.467	0.795
7	SgreTa0005600	966.179		279.788		2.927	819.045	0.795
7	SgreTa0013453	9.930		8.482		61.614	522.632	0.794
7	SgreTa0008219	145.183		143.957		0.797	114.758	0.794
7	SgreTa0006308	8.248		9.049		0.111	1.008	0.303
7	SgreTa0001661	4.515		3.163		38.206	120.862	0.754
7	SgreTa0014626	8.443		6.388		160.490	1025.240	0.791
8	SgreTa0007432		1.189		1.296	9.430	7.276	0.395
8	SgreTa0001469	4.905		3.109		89.858	279.396	0.749
8	SgreTa0007802	2052.856		652.039		1.063	693.338	0.788
8	SgreTa0005616	552.396		234.275		0.489	114.514	0.788
8	SgreTa0017664	26.449		7.225		7.118	51.426	0.785
8	SgreTa0009118		8.144		9.556	82.176	8.599	0.787
8	SgreTa0000088		1.443		1.496	34.508	23.071	0.526
8	SgreTa0014975	3.818		2.915		18.201	53.056	0.710

8	SgreTb0019973	29.591		20.892		55.364	1156.651	0.788
8	SgreTb0006243	537.340		297.963		2.284	680.550	0.788
8	SgreTa0017707	344.572		219.619		16.926	3717.263	0.788
8	SgreTa0017736	20.254		11.347		1.017	11.544	0.763
8	SgreTa0007897	116.944		69.794		8.998	628.012	0.788
8	SgreTa0008528	420.588		251.695		3.422	861.195	0.788
8	SgreTa0005600	318.652		179.306		7.170	1285.584	0.788
8	SgreTa0013453	5.827		5.301		73.218	388.135	0.780
8	SgreTa0008219	133.217		91.006		1.483	134.976	0.788
8	SgreTd0008886	1.776			2.530	1.933	0.764	0.467
8	SgreTd0014875	1.260		1.143		10.234	11.697	0.345
8	SgreTa0006386		6.385		23.445	18.820	0.803	0.779
8	SgreTa0006977		5.063		13.209	9.183	0.695	0.750
8	SgreTa0006308	2.995			1.383	5.007	3.620	0.390
8	SgreTa0002186	820.939		463.758		2.510	1164.179	0.788
8	SgreTa0001661	5.725		4.409		42.609	187.852	0.771
8	SgreTb0016047	5.226		4.234		30.761	130.242	0.771
8	SgreTb0016047	5.048		4.234		30.761	130.242	0.771
8	SgreTa0014626	7.975		4.671		199.900	933.699	0.773
8	SgreTa0008504	8.862		6.990		76.345	533.636	0.786
10	SgreTa0007432		1.999		1.396	9.752	6.988	0.420
10	SgreTa0001469	5.781		5.850		109.290	639.310	0.773
10	SgreTa0007802	132.766		250.286		2.403	601.355	0.781
10	SgreTa0005616	64.339		58.145		0.902	52.426	0.780
10	SgreTa0011044	not detected				filtered out		
10	SgreTa0006252	1.805		3.043		52.999	161.261	0.741
10	SgreTa0017664		9.079		3.391	166.597	49.135	0.742
10	SgreTa0005054		2.158		1.237	34.081	27.554	0.396
10	SgreTa0002027		DOWN indefinitely		87.018	3.994	0.046	0.642
10	SgreTa0009118		45.352		33.703	69.800	2.071	0.781
10	SgreTa0000088		11.293		9.208	43.460	4.720	0.779
10	SgreTa0014975	3.435		4.368		11.542	50.413	0.762
10	SgreTa0001826	20.587		24.972		11.469	286.400	0.781
10	SgreTa0000488	12.479		34.903		9.829	343.044	0.781
10	SgreTa0009559	29.148		56.862		5.898	335.386	0.781
10	SgreTa0003305	3.697		5.089		42.278	215.172	0.773
10	SgreTd0003949	14.818		20.871		2.640	55.104	0.780
10	SgreTb0006243	87.476		113.407		5.083	576.448	0.781
10	SgreTa0017707	1001.576		756.230		9.643	7292.232	0.919

10	SgreTa0017736	22.014		47.457		1.186	56.284	0.780
10	SgreTa0007897	103.446		136.655		48.453	6621.339	0.781
10	SgreTa0008528	180.946		202.895		8.199	1663.469	0.781
10	SgreTa0001449	UP indefinitely		582.222		0.369	214.943	0.919
10	SgreTa0005600	56.826		62.483		12.381	773.605	0.781
10	SgreTc0000004	UP indefinitely		2844.050		0.154	439.013	0.998
10	SgreTc0000003	3.309		4.751		2.622	12.459	0.740
10	SgreTc0000003	3.687		4.751		2.622	12.459	0.740
10	SgreTa0013453	2.349		3.157		82.567	260.688	0.741
10	SgreTa0008219	36.004		37.305		4.463	166.492	0.781
10	SgreTa0008497	63.642		55.465		2.680	148.627	0.781
10	SgreTa0002186	225.140		195.257		2.298	448.716	0.781
10	SgreTa0001661	3.569		4.482		34.555	154.888	0.764
10	SgreTb0016047	6.146		6.000		21.481	128.879	0.777
10	SgreTa0014626	3.451		5.045		167.728	846.249	0.773
11	SgreTa0007432		1.320		1.732	10.454	6.036	0.514
11	SgreTa0001469	5.808		5.190		117.377	609.215	0.769
11	SgreTa0007802	100.910		143.963		2.751	396.066	0.776
11	SgreTa0005616	18.886		58.723		0.980	57.520	0.775
11	SgreTa0017664	1256.088		12.672		55.939	708.860	0.776
11	SgreTa0014975	4.698		4.449		14.202	63.189	0.758
11	SgreTb0006243	57.231		108.492		5.052	548.138	0.776
11	SgreTa0017707	1216.859		358.692		18.704	6708.830	0.776
11	SgreTa0017736	133.453		210.082		1.312	275.635	0.776
11	SgreTa0007897	428.417		352.913		48.413	17085.542	0.776
11	SgreTa0008528	78.880		73.748		12.770	941.764	0.776
11	SgreTa0005600	79.572		68.596		14.671	1006.405	0.776
11	SgreTa0013453	3.110		2.931		85.577	250.815	0.699
11	SgreTa0008219	19.265		24.288		4.807	116.746	0.776
11	SgreTa0001661	4.935		4.377		34.422	150.654	0.759
11	SgreTa0014626	4.514		4.126		161.557	666.631	0.759
12	SgreTa0007432	1.306		1.009		5.591	5.640	0.368
12	SgreTa0001469	10.590		6.115		132.137	808.004	0.754
12	SgreTa0007802	30.338		19.349		12.445	240.795	0.756
12	SgreTa0005616	13.108		3.231		16.053	51.872	0.720
12	SgreTa0017664	2689.973		460.128		15.436	7102.347	0.756
12	SgreTb0006243	11.666		10.639		40.743	433.469	0.756
12	SgreTa0017707	4391.195		743.840		8.249	6135.875	0.916
12	SgreTa0017736	964.373		291.207		1.516	441.598	0.756
12	SgreTa0007897	2008.506		615.488		31.925	19649.559	0.916

12	SgreTa0008528	506.989	179.902	4.855	873.417	0.756
12	SgreTa0005600	56.899	56.893	18.258	1038.782	0.756
12	SgreTa0013453	2.626	1.967	126.721	249.303	0.575
12	SgreTa0008219	15.694	8.224	16.314	134.174	0.756
12	SgreTa0001661	6.011	3.578	39.526	141.416	0.725
12	SgreTa0014626	3.395	2.522	242.996	612.785	0.681
13	SgreTa0007432	1.771	1.474	3.521	5.189	0.431
13	SgreTa0001469	10.436	5.266	136.452	718.578	0.732
13	SgreTa0007802	18.212	15.472	6.595	102.045	0.737
13	SgreTa0005616	7.423	2.137	14.119	30.177	0.641
13	SgreTa0017664	1748.639	576.461	26.670	15373.994	0.911
13	SgreTb0006243	46.612	47.622	6.077	289.402	0.737
13	SgreTa0017707	8527.308	469.508	7.262	3409.772	0.911
13	SgreTa0017736	1344.749	658.489	0.672	442.276	0.911
13	SgreTa0007897	1243.649	456.919	37.460	17116.002	0.911
13	SgreTa0008528	UP indefinitely	433.678	2.621	1136.614	0.738
13	SgreTa0005600	83.004	72.572	15.901	1153.927	0.738
13	SgreTa0013453	2.524	1.873	112.755	211.156	0.565
13	SgreTa0008219	29.562	10.248	13.991	143.386	0.737
13	SgreTa0001661	14.842	5.235	33.357	174.626	0.731
13	SgreTa0014626	5.341	3.954	164.431	650.184	0.711

[a] "UP indefinitely": not detected in the legs after 35 cycles, "DOWN indefinitely": not detected in the pleuropodia; compare with the low RPKM in LEG and PLP samples, respectively

[b] significant upregulation (UP) or downregulation (DOWN) (fold change between expression in pleuropodia and legs) are highlighted in magenta and blue, respectively (thresholds: prob > 0.7, RPKM > 10, fold change > 2; prob below threshold highlighted in grey)

[c] not applicable - expression too low

Supplementary Table 6. GOs enriched in the downregulated DEGs from the highly secreting pleuropodia (joined sample 10, 11 and 12 days) – First 100 terms are shown.

category	over represented pvalue	num DEInCat	num InCat	term	ontology[a]	over represented FDR
GO:0048856	2.00E-20	320	2000	anatomical structure development	BP	3.36E-16
GO:0051301	1.75E-19	88	292	cell division	BP	1.47E-15
GO:0007010	1.56E-17	112	446	cytoskeleton organization	BP	6.54E-14
GO:0022402	2.10E-16	126	567	cell cycle process	BP	5.89E-13
GO:0031032	1.25E-15	37	81	actomyosin structure organization	BP	3.00E-12
GO:0007049	5.33E-15	81	301	cell cycle	BP	1.12E-11
GO:0044767	1.21E-14	375	2646	single-organism developmental process	BP	2.26E-11
GO:0051276	1.35E-14	61	195	chromosome organization	BP	2.27E-11
GO:0048513	3.75E-14	151	838	animal organ development	BP	5.72E-11
GO:0032502	1.16E-13	389	2822	developmental process	BP	1.62E-10
GO:1903047	4.45E-13	90	395	mitotic cell cycle process	BP	5.49E-10
GO:0009888	4.57E-13	80	348	tissue development	BP	5.49E-10
GO:0071840	7.41E-13	380	2732	cellular component organization or biogenesis	BP	8.31E-10
GO:0016043	1.26E-12	374	2689	cellular component organization	BP	1.25E-09
GO:0045214	1.60E-12	25	51	sarcomere organization	BP	1.49E-09
GO:0022414	2.36E-12	169	999	reproductive process	BP	2.08E-09
GO:0097435	1.50E-11	58	217	supramolecular fiber organization	BP	1.27E-08
GO:0071103	3.16E-11	28	64	DNA conformation change	BP	2.53E-08
GO:0007017	1.35E-10	79	377	microtubule-based process	BP	1.03E-07
GO:0000226	1.51E-10	54	202	microtubule cytoskeleton organization	BP	1.10E-07
GO:0006996	1.99E-10	196	1224	organelle organization	BP	1.39E-07
GO:0006323	3.62E-10	21	40	DNA packaging	BP	2.44E-07
GO:0006260	3.91E-10	37	114	DNA replication	BP	2.53E-07
GO:0030261	1.55E-09	20	39	chromosome condensation	BP	7.92E-07
GO:0000278	1.70E-09	30	83	mitotic cell cycle	BP	8.42E-07
GO:0035295	3.63E-09	47	187	tube development	BP	1.69E-06
GO:0007444	5.94E-09	29	89	imaginal disc development	BP	2.56E-06
GO:0009653	1.39E-08	157	1030	anatomical structure morphogenesis	BP	5.72E-06
GO:0007517	2.61E-08	25	70	muscle organ development	BP	1.04E-05
GO:0042127	3.83E-08	88	474	regulation of cell proliferation	BP	1.43E-05
GO:0010564	5.03E-08	69	338	regulation of cell cycle process	BP	1.76E-05
GO:0030036	6.24E-08	45	175	actin cytoskeleton organization	BP	2.14E-05
GO:1903046	7.39E-08	33	113	meiotic cell cycle process	BP	2.49E-05
GO:0032501	1.53E-07	240	1835	multicellular organismal process	BP	5.04E-05
GO:0051726	1.63E-07	95	533	regulation of cell cycle	BP	5.28E-05
GO:0050793	1.67E-07	153	989	regulation of developmental process	BP	5.28E-05

GO:0009886	2.02E-07	44	182	post-embryonic animal morphogenesis	BP	6.28E-05
GO:0030029	3.27E-07	46	193	actin filament-based process	BP	9.37E-05
GO:0006270	3.29E-07	14	28	DNA replication initiation	BP	9.37E-05
GO:0044702	3.52E-07	123	766	single organism reproductive process	BP	9.87E-05
GO:1901990	4.50E-07	41	165	regulation of mitotic cell cycle phase transition	BP	0.000122089
GO:0030703	4.60E-07	9	12	eggshell formation	BP	0.00012278
GO:0044699	4.86E-07	678	6032	single-organism process	BP	0.000127791
GO:0060429	5.28E-07	39	161	epithelium development	BP	0.000136482
GO:0006275	5.68E-07	17	41	regulation of DNA replication	BP	0.000138504
GO:0007498	6.68E-07	18	46	mesoderm development	BP	0.00015832
GO:0051783	8.24E-07	42	174	regulation of nuclear division	BP	0.000192552
GO:0007346	1.00E-06	65	338	regulation of mitotic cell cycle	BP	0.000231045
GO:0007088	1.20E-06	40	163	regulation of mitotic nuclear division	BP	0.000262458
GO:0044707	1.25E-06	209	1578	single-multicellular organism process	BP	0.000270362
GO:1901987	1.28E-06	42	176	regulation of cell cycle phase transition	BP	0.000272555
GO:0007076	1.36E-06	12	22	mitotic chromosome condensation	BP	0.00028512
GO:2000026	1.61E-06	115	715	regulation of multicellular organismal development	BP	0.000330645
GO:0090068	1.82E-06	34	128	positive regulation of cell cycle process	BP	0.000368697
GO:0051239	2.61E-06	153	1042	regulation of multicellular organismal process	BP	0.000521508
GO:0061077	2.80E-06	13	33	chaperone-mediated protein folding	BP	0.000547652
GO:0007304	2.84E-06	8	11	chorion-containing eggshell formation	BP	0.000549565
GO:0035220	3.44E-06	19	59	wing disc development	BP	0.000646364
GO:0048869	3.46E-06	201	1476	cellular developmental process	BP	0.000646364
GO:0032989	3.86E-06	74	451	cellular component morphogenesis	BP	0.000704622
GO:0043062	4.36E-06	25	90	extracellular structure organization	BP	0.000779886
GO:0090329	4.91E-06	11	23	regulation of DNA-dependent DNA replication	BP	0.000868797
GO:0042559	5.27E-06	8	14	pteridine-containing compound biosynthetic process	BP	0.000923292
GO:0002066	6.43E-06	22	76	columnar/cuboidal epithelial cell development	BP	0.001114131
GO:0022412	6.67E-06	72	405	cellular process involved in reproduction in multicellular organism	BP	0.001145056
GO:0060052	9.62E-06	7	10	neurofilament cytoskeleton organization	BP	0.001586311
GO:0007552	1.05E-05	12	40	metamorphosis	BP	0.001685364
GO:0051983	1.10E-05	19	58	regulation of chromosome segregation	BP	0.00175249
GO:0035114	1.13E-05	24	88	imaginal disc-derived appendage morphogenesis	BP	0.001777181
GO:0007015	1.19E-05	30	122	actin filament organization	BP	0.001832454

GO:0000904	1.26E-05	23	84	cell morphogenesis involved in differentiation	BP	0.001926008
GO:0045297	1.34E-05	8	25	post-mating behavior	BP	0.002030503
GO:0061061	1.59E-05	33	162	muscle structure development	BP	0.002373115
GO:0048646	1.61E-05	74	433	anatomical structure formation involved in morphogenesis	BP	0.002378056
GO:0002064	1.77E-05	27	108	epithelial cell development	BP	0.002539527
GO:0042558	1.77E-05	10	23	pteridine-containing compound metabolic process	BP	0.002539527
GO:0006281	1.90E-05	60	335	DNA repair	BP	0.002689969
GO:0030071	2.12E-05	15	41	regulation of mitotic metaphase/ anaphase transition	BP	0.002893062
GO:1902099	2.12E-05	15	41	regulation of metaphase/ anaphase transition of cell cycle	BP	0.002893062
GO:0030198	2.57E-05	22	82	extracellular matrix organization	BP	0.003437198
GO:0044763	2.58E-05	548	4818	single-organism cellular process	BP	0.003437198
GO:0030707	2.60E-05	19	63	ovarian follicle cell development	BP	0.003439415
GO:0007527	2.74E-05	7	10	adult somatic muscle development	BP	0.003439415
GO:0045168	2.74E-05	19	65	cell-cell signaling involved in cell fate commitment	BP	0.003439415
GO:0046331	2.74E-05	19	65	lateral inhibition	BP	0.003439415
GO:0045841	2.74E-05	10	20	negative regulation of mitotic metaphase/anaphase transition	BP	0.003439415
GO:1902100	2.74E-05	10	20	negative regulation of metaphase/anaphase transition of cell cycle	BP	0.003439415
GO:1905819	2.74E-05	10	20	negative regulation of chromosome separation	BP	0.003439415
GO:2000816	2.74E-05	10	20	negative regulation of mitotic sister chromatid separation	BP	0.003439415
GO:0048609	2.90E-05	68	409	multicellular organismal reproductive process	BP	0.003612766
GO:0033045	3.02E-05	17	52	regulation of sister chromatid segregation	BP	0.003706573
GO:0010965	3.11E-05	15	42	regulation of mitotic sister chromatid separation	BP	0.0037589
GO:1905818	3.11E-05	15	42	regulation of chromosome separation	BP	0.0037589
GO:0007519	3.20E-05	11	24	skeletal muscle tissue development	BP	0.003837808
GO:0032467	3.41E-05	8	13	positive regulation of cytokinesis	BP	0.003984256
GO:0042335	3.67E-05	20	78	cuticle development	BP	0.004254892
GO:0051253	3.69E-05	77	460	negative regulation of RNA metabolic process	BP	0.004254892
GO:0035120	3.97E-05	22	82	post-embryonic appendage morphogenesis	BP	0.004537709
GO:0033046	4.09E-05	11	25	negative regulation of sister chromatid segregation	BP	0.004610621
GO:0051985	4.09E-05	11	25	negative regulation of chromosome segregation	BP	0.004610621

[a] BP, biological process; CC, cellular component; MF, molecular function

Supplementary Table 7. GOs enriched in the upregulated DEGs from the highly secreting pleuropodia (joined sample 10, 11 and 12 days).

category	over represented pvalue	num DEInCat	num InCat	term	ontology[a]	over represented FDR
GO:0006811	1.47E-19	77	534	ion transport	BP	1.24E-15
GO:0034220	2.23E-16	43	221	ion transmembrane transport	BP	4.69E-13
GO:0090662	7.94E-16	16	24	ATP hydrolysis coupled transmembrane transport	BP	1.48E-12
GO:0015672	4.65E-15	33	149	monovalent inorganic cation transport	BP	6.76E-12
GO:0055085	4.82E-15	56	381	transmembrane transport	BP	6.76E-12
GO:0015988	6.36E-15	14	19	energy coupled proton transmembrane transport, against electrochemical gradient	BP	7.64E-12
GO:0015991	6.36E-15	14	19	ATP hydrolysis coupled proton transport	BP	7.64E-12
GO:0099131	4.55E-14	14	21	ATP hydrolysis coupled ion transmembrane transport	BP	4.50E-11
GO:0099132	4.55E-14	14	21	ATP hydrolysis coupled cation transmembrane transport	BP	4.50E-11
GO:0006820	1.20E-13	43	253	anion transport	BP	1.06E-10
GO:0006818	8.74E-13	17	41	hydrogen transport	BP	7.00E-10
GO:0015711	1.50E-12	37	207	organic anion transport	BP	1.14E-09
GO:0015992	7.45E-12	16	40	proton transport	BP	5.45E-09
GO:0044765	8.20E-11	92	961	single-organism transport	BP	4.90E-08
GO:0007311	1.21E-10	12	24	maternal specification of dorsal/ventral axis, oocyte, germ-line encoded	BP	6.77E-08
GO:0006812	1.34E-10	41	302	cation transport	BP	7.26E-08
GO:1902600	2.07E-10	14	35	hydrogen ion transmembrane transport	BP	1.06E-07
GO:1902578	6.71E-10	95	1043	single-organism localization	BP	3.05E-07
GO:0098655	1.92E-09	24	133	cation transmembrane transport	BP	7.67E-07
GO:0008063	7.06E-09	15	50	Toll signaling pathway	BP	2.58E-06
GO:1901615	1.75E-08	36	273	organic hydroxy compound metabolic process	BP	5.65E-06
GO:0007310	1.82E-08	13	39	oocyte dorsal/ventral axis specification	BP	5.76E-06
GO:0098660	2.46E-08	24	145	inorganic ion transmembrane transport	BP	7.34E-06
GO:0098662	3.09E-08	21	117	inorganic cation transmembrane transport	BP	8.80E-06
GO:0009950	4.22E-08	15	55	dorsal/ventral axis specification	BP	1.14E-05
GO:0015849	5.06E-08	23	132	organic acid transport	BP	1.31E-05
GO:0046942	5.06E-08	23	132	carboxylic acid transport	BP	1.31E-05
GO:0006865	6.50E-08	17	76	amino acid transport	BP	1.58E-05
GO:0003333	1.02E-07	12	37	amino acid transmembrane transport	BP	2.42E-05
GO:0006629	1.29E-07	64	687	lipid metabolic process	BP	3.02E-05

GO:0007309	1.69E-07	14	52	oocyte axis specification	BP	3.80E-05
GO:0006814	3.19E-07	15	73	sodium ion transport	BP	6.96E-05
GO:0007370	4.49E-07	8	17	ventral furrow formation	BP	9.44E-05
GO:0006809	5.39E-07	5	7	nitric oxide biosynthetic process	BP	0.000109252
GO:0046209	5.39E-07	5	7	nitric oxide metabolic process	BP	0.000109252
GO:0044281	1.22E-06	85	1048	small molecule metabolic process	BP	0.000226884
GO:0044710	1.23E-06	144	2090	single-organism metabolic process	BP	0.000226884
GO:1903825	1.91E-06	12	47	organic acid transmembrane transport	BP	0.00032152
GO:1905039	1.91E-06	12	47	carboxylic acid transmembrane transport	BP	0.00032152
GO:0098656	2.65E-06	15	75	anion transmembrane transport	BP	0.000423703
GO:0006810	4.16E-06	136	1930	transport	BP	0.000641496
GO:0044699	4.20E-06	346	6032	single-organism process	BP	0.000642071
GO:0051234	4.73E-06	139	1983	establishment of localization	BP	0.000714453
GO:0006885	4.76E-06	10	35	regulation of pH	BP	0.000714453
GO:0044703	5.37E-06	13	67	multi-organism reproductive process	BP	0.00079928
GO:2001057	6.39E-06	5	9	reactive nitrogen species metabolic process	BP	0.000925751
GO:0006026	8.56E-06	11	47	aminoglycan catabolic process	BP	0.001199491
GO:0055067	8.96E-06	10	37	monovalent inorganic cation homeostasis	BP	0.001245784
GO:1903409	1.93E-05	5	11	reactive oxygen species biosynthetic process	BP	0.002598165
GO:0051453	2.03E-05	9	33	regulation of intracellular pH	BP	0.002715202
GO:0046348	2.30E-05	7	19	amino sugar catabolic process	BP	0.003041764
GO:0030641	2.74E-05	9	34	regulation of cellular pH	BP	0.003577987
GO:0045851	2.88E-05	6	15	pH reduction	BP	0.003727792
GO:1901136	3.44E-05	15	93	carbohydrate derivative catabolic process	BP	0.004385845
GO:0050801	3.60E-05	26	237	ion homeostasis	BP	0.00454847
GO:0030004	3.68E-05	9	35	cellular monovalent inorganic cation homeostasis	BP	0.004619921
GO:0042940	3.92E-05	4	6	D-amino acid transport	BP	0.004883535
GO:0006869	4.07E-05	18	130	lipid transport	BP	0.004979036
GO:0019835	4.09E-05	5	13	cytolysis	BP	0.004979036
GO:0005975	4.52E-05	35	364	carbohydrate metabolic process	BP	0.005466278
GO:0006582	5.02E-05	10	47	melanin metabolic process	BP	0.005988923
GO:0007035	5.38E-05	5	10	vacuolar acidification	BP	0.006373312
GO:0009617	5.70E-05	17	123	response to bacterium	BP	0.006708191
GO:0048878	5.84E-05	33	341	chemical homeostasis	BP	0.006773732
GO:0043207	5.84E-05	33	369	response to external biotic stimulus	BP	0.006773732
GO:1901072	7.99E-05	6	16	glucosamine-containing compound catabolic process	BP	0.008951938
GO:0072593	8.36E-05	9	43	reactive oxygen species metabolic process	BP	0.009311797

GO:0009607	9.01E-05	33	377	response to biotic stimulus	BP	0.009925828
GO:0035006	9.47E-05	8	32	melanization defense response	BP	0.010268044
GO:0055088	0.0001012	12	70	lipid homeostasis	BP	0.010843485
GO:0051179	0.0001095	148	2272	localization	BP	0.011649984
GO:0044706	0.0001168	10	55	multi-multicellular organism process	BP	0.012352242
GO:0050830	0.0001252	9	43	defense response to Gram-positive bacterium	BP	0.013158789
GO:0009798	0.0001292	16	109	axis specification	BP	0.013446589
GO:0030001	0.0001295	23	222	metal ion transport	BP	0.013446589
GO:0009620	0.0001398	9	42	response to fungus	BP	0.014424836
GO:0009605	0.0001444	51	666	response to external stimulus	BP	0.014716665
GO:0006003	0.0001636	3	3	fructose 2,6-bisphosphate metabolic process	BP	0.016276968
GO:0006665	0.0001704	11	61	sphingolipid metabolic process	BP	0.016755635
GO:0051704	0.0001748	44	549	multi-organism process	BP	0.016992012
GO:0018958	0.0001763	13	88	phenol-containing compound metabolic process	BP	0.017039357
GO:0051707	0.0001907	28	311	response to other organism	BP	0.018223573
GO:0065008	0.0001928	103	1526	regulation of biological quality	BP	0.018318887
GO:0006066	0.0002057	20	175	alcohol metabolic process	BP	0.019211887
GO:0051452	0.0002413	5	14	intracellular pH reduction	BP	0.02229664
GO:0006563	0.0002432	4	7	L-serine metabolic process	BP	0.022346163
GO:0010817	0.0003176	22	213	regulation of hormone levels	BP	0.028404042
GO:0071825	0.0003465	6	22	protein-lipid complex subunit organization	BP	0.030502766
GO:0071827	0.0003465	6	22	plasma lipoprotein particle organization	BP	0.030502766
GO:0006032	0.0004406	5	14	chitin catabolic process	BP	0.03838474
GO:0034368	0.0004595	5	14	protein-lipid complex remodeling	BP	0.039247694
GO:0034369	0.0004595	5	14	plasma lipoprotein particle remodeling	BP	0.039247694
GO:0034375	0.0004595	5	14	high-density lipoprotein particle remodeling	BP	0.039247694
GO:0042742	0.0004629	14	109	defense response to bacterium	BP	0.039308893
GO:0034374	0.0005063	4	9	low-density lipoprotein particle remodeling	BP	0.042777281
GO:0045087	0.0005142	18	162	innate immune response	BP	0.043231252
GO:0019752	0.0005449	43	513	carboxylic acid metabolic process	BP	0.04558214
GO:0006564	0.0005656	3	4	L-serine biosynthetic process	BP	0.047084223
GO:0032367	0.0005741	5	18	intracellular cholesterol transport	BP	0.047557237
GO:0032787	0.0005854	26	263	monocarboxylic acid metabolic process	BP	0.048015835

[a] BP, biological process; CC, cellular component; MF, molecular function

Supplementary Table 8. GOs enriched in the downregulated DEGs from each developmental stage (FDR < e-5); Only the first 10 GO terms of each time-point are shown.

day	category	over represented pvalue	num DEInCat	num InCat	term	ontology[a]	over represented FDR
7	GO:0051301	1.50E-11	23	292	cell division	BP	2.52E-07
	GO:0007049	1.34E-09	21	301	cell cycle	BP	1.13E-05
8	GO:0051301	2.76E-11	28	292	cell division	BP	4.65E-07
	GO:0005488	2.99E-10	185	7055	binding	MF	1.76E-06
	GO:0007049	3.14E-10	27	301	cell cycle	BP	1.76E-06
	GO:0005832	2.00E-09	6	8	chaperonin-containing T-complex	CC	6.33E-06
	GO:0101031	2.00E-09	6	8	chaperone complex	CC	6.33E-06
	GO:0016043	2.55E-09	93	2689	cellular component organization	BP	6.33E-06
	GO:0071840	2.63E-09	94	2732	cellular component organization or biogenesis	BP	6.33E-06
	GO:0005634	1.30E-08	95	2897	nucleus	CC	2.74E-05
	GO:0097159	3.75E-08	117	4057	organic cyclic compound binding	MF	7.00E-05
	GO:0042559	5.81E-08	6	14	pteridine-containing compound biosynthetic process	BP	9.76E-05
8-9	GO:0051301	9.46E-16	32	292	cell division	BP	1.59E-11
	GO:0005634	1.43E-13	100	2897	nucleus	CC	1.20E-09
	GO:0005488	1.18E-11	172	7055	binding	MF	6.62E-08
	GO:0007049	2.34E-11	27	301	cell cycle	BP	9.82E-08
	GO:0044427	1.03E-09	30	473	chromosomal part	CC	3.46E-06
	GO:0071103	2.76E-09	12	64	DNA conformation change	BP	7.74E-06
	GO:0048856	5.90E-09	68	2000	anatomical structure development	BP	1.32E-05
	GO:0051276	6.93E-09	19	195	chromosome organization	BP	1.32E-05
	GO:0071840	7.43E-09	85	2732	cellular component organization or biogenesis	BP	1.32E-05
	GO:0016043	7.83E-09	84	2689	cellular component organization	BP	1.32E-05
10	GO:0005198	2.07E-15	45	502	structural molecule activity	MF	3.47E-11
	GO:0042302	1.47E-14	20	113	structural constituent of cuticle	MF	1.08E-10
	GO:0071103	2.56E-14	19	64	DNA conformation change	BP	1.08E-10
	GO:0006323	2.58E-14	16	40	DNA packaging	BP	1.08E-10
	GO:0044427	1.57E-13	46	473	chromosomal part	CC	5.26E-10
	GO:0051276	1.95E-13	30	195	chromosome organization	BP	5.47E-10
	GO:0030261	3.22E-13	15	39	chromosome condensation	BP	7.73E-10
	GO:0051301	5.77E-13	36	292	cell division	BP	1.21E-09
	GO:0071840	1.07E-12	135	2732	cellular component organization or biogenesis	BP	2.01E-09
	GO:0016043	1.58E-12	133	2689	cellular component organization	BP	2.66E-09
11	GO:0071840	5.54E-16	153	2732	cellular component organization or biogenesis	BP	9.32E-12
	GO:0016043	1.73E-15	150	2689	cellular component organization	BP	1.46E-11
	GO:0044427	3.31E-15	51	473	chromosomal part	CC	1.85E-11

	GO:0051276	5.27E-14	32	195	chromosome organization	BP	2.21E-10
	GO:0051301	7.92E-14	39	292	cell division	BP	2.67E-10
	GO:0071103	1.22E-13	19	64	DNA conformation change	BP	3.41E-10
	GO:0005488	1.45E-13	284	7055	binding	MF	3.41E-10
	GO:0005634	1.62E-13	151	2897	nucleus	CC	3.41E-10
	GO:0007049	2.49E-13	39	301	cell cycle	BP	4.66E-10
	GO:0044428	1.64E-12	121	2074	nuclear part	CC	2.69E-09
12	GO:0005198	1.10E-13	46	502	structural molecule activity	MF	1.86E-09
	GO:0031032	1.16E-12	20	81	actomyosin structure organization	BP	9.74E-09
	GO:0044183	1.78E-12	9	11	protein binding involved in protein folding	MF	9.96E-09
	GO:0097435	2.18E-11	30	217	supramolecular fiber organization	BP	9.15E-08
	GO:0005832	2.49E-10	7	8	chaperonin-containing T-complex	CC	6.62E-07
	GO:0101031	2.49E-10	7	8	chaperone complex	CC	6.62E-07
	GO:0044449	2.77E-10	22	124	contractile fiber part	CC	6.62E-07
	GO:0042302	3.29E-10	17	113	structural constituent of cuticle	MF	6.62E-07
	GO:0045214	3.54E-10	14	51	sarcomere organization	BP	6.62E-07
	GO:0006457	3.74E-09	19	115	protein folding	BP	6.28E-06
	GO:0030036	5.70E-09	24	175	actin cytoskeleton organization	BP	8.71E-06
13	GO:0042302	1.77E-69	61	113	structural constituent of cuticle	MF	2.98E-65
	GO:0005198	3.79E-46	89	502	structural molecule activity	MF	3.19E-42
	GO:0031032	2.37E-11	19	81	actomyosin structure organization	BP	1.33E-07
	GO:0006091	9.91E-11	22	132	generation of precursor metabolites and energy	BP	3.91E-07
	GO:0046034	1.16E-10	16	68	ATP metabolic process	BP	3.91E-07
	GO:0044449	2.74E-10	24	124	contractile fiber part	CC	7.68E-07
	GO:0044455	4.56E-10	20	124	mitochondrial membrane part	CC	1.09E-06
	GO:0045214	1.02E-09	14	51	sarcomere organization	BP	2.15E-06
	GO:0006090	1.40E-09	12	38	pyruvate metabolic process	BP	2.62E-06
	GO:0009205	2.40E-09	16	82	purine ribonucleoside triphosphate metabolic process	BP	4.04E-06

[a] BP, biological process; CC, cellular component; MF, molecular function

Supplementary Table 9. GOs enriched in the upregulated DEGs from each developmental stage (FDR < e-5); Only the first 10 GO terms of each time-point are shown.

day	category	over represented pvalue	num DEInCat	num InCat	term	ontology[a]	over represented FDR
4							
5	GO:0044420	4.846E-09	8	66	extracellular matrix component	CC	8.14836E-05
	GO:0005604	9.752E-09	7	46	basement membrane	CC	8.19932E-05
6	GO:0006030	1.032E-13	11	26	chitin metabolic process	BP	1.73589E-09
	GO:1901071	1.561E-11	11	37	glucosamine-containing compound metabolic process	BP	1.31253E-07
	GO:0006040	2.608E-11	12	50	amino sugar metabolic process	BP	1.46188E-07
	GO:0044421	2.251E-10	51	1255	extracellular region part	CC	9.46145E-07
7	GO:0090662	7.616E-20	16	24	ATP hydrolysis coupled transmembrane transport	BP	1.28067E-15
	GO:0015988	2.554E-18	14	19	energy coupled proton transmembrane transport, against electrochemical gradient	BP	1.43142E-14
	GO:0015991	2.554E-18	14	19	ATP hydrolysis coupled proton transport	BP	1.43142E-14
	GO:0099131	1.415E-17	14	21	ATP hydrolysis coupled ion transmembrane transport	BP	4.75697E-14
	GO:0099132	1.415E-17	14	21	ATP hydrolysis coupled cation transmembrane transport	BP	4.75697E-14
	GO:0044425	1.846E-17	153	3119	membrane part	CC	5.17435E-14
	GO:0016021	3.583E-16	124	2321	integral component of membrane	CC	8.60757E-13
	GO:0006818	7.457E-16	16	41	NA	NA	1.56745E-12
	GO:0031224	1.117E-15	125	2397	intrinsic component of membrane	CC	2.08631E-12
	GO:0015992	1.137E-14	15	40	NA	NA	1.91225E-11
8	GO:0090662	4.98E-20	1.60E+01	2.40E+01	ATP hydrolysis coupled transmembrane transport	BP	8.37E-16
	GO:0015988	1.49E-18	1.40E+01	1.90E+01	energy coupled proton transmembrane transport, against electrochemical gradient	BP	8.37E-15
	GO:0015991	1.49E-18	1.40E+01	1.90E+01	ATP hydrolysis coupled proton transport	BP	8.37E-15
	GO:0099131	9.69E-18	1.40E+01	2.10E+01	ATP hydrolysis coupled ion transmembrane transport	BP	3.26E-14
	GO:0099132	9.69E-18	1.40E+01	2.10E+01	ATP hydrolysis coupled cation transmembrane transport	BP	3.26E-14
	GO:0044425	5.63E-17	1.50E+02	3.12E+03	membrane part	CC	1.58E-13
	GO:0006818	7.48E-16	1.60E+01	4.10E+01	NA	NA	1.80E-12
	GO:0044710	1.06E-15	1.13E+02	2.09E+03	NA	NA	2.23E-12
	GO:0015992	1.15E-14	1.50E+01	4.00E+01	NA	NA	2.14E-11
	GO:0016021	2.10E-14	1.18E+02	2.32E+03	integral component of membrane	CC	3.54E-11
8-9	GO:0090662	1.72E-19	1.60E+01	2.40E+01	ATP hydrolysis coupled transmembrane transport	BP	2.88E-15

	GO:0015988	4.61E-18	1.40E+01	1.90E+01	energy coupled proton transmembrane transport, against electrochemical gradient	BP	2.58E-14
	GO:0015991	4.61E-18	1.40E+01	1.90E+01	ATP hydrolysis coupled proton transport	BP	2.58E-14
	GO:0099131	2.86E-17	1.40E+01	2.10E+01	ATP hydrolysis coupled ion transmembrane transport	BP	9.61E-14
	GO:0099132	2.86E-17	1.40E+01	2.10E+01	ATP hydrolysis coupled cation transmembrane transport	BP	9.61E-14
	GO:0015992	3.20E-14	1.50E+01	4.00E+01	NA	NA	8.96E-11
	GO:0006818	5.28E-14	1.50E+01	4.10E+01	NA	NA	1.27E-10
	GO:1902600	1.05E-13	1.40E+01	3.50E+01	proton transmembrane transport	BP	2.20E-10
	GO:0046961	6.80E-13	9.00E+00	1.10E+01	proton-transporting ATPase activity, rotational mechanism	MF	1.27E-09
	GO:0044425	1.61E-12	1.46E+02	3.12E+03	membrane part	CC	2.67E-09
10	GO:0090662	4.15E-19	1.60E+01	2.40E+01	ATP hydrolysis coupled transmembrane transport	BP	4.65E-15
	GO:0006811	5.54E-19	5.80E+01	5.34E+02	ion transport	BP	4.65E-15
	GO:0044425	4.73E-18	1.68E+02	3.12E+03	membrane part	CC	2.65E-14
	GO:0031224	7.66E-18	1.41E+02	2.40E+03	intrinsic component of membrane	CC	2.81E-14
	GO:0015988	1.00E-17	1.40E+01	1.90E+01	energy coupled proton transmembrane transport, against electrochemical gradient	BP	2.81E-14
	GO:0015991	1.00E-17	1.40E+01	1.90E+01	ATP hydrolysis coupled proton transport	BP	2.81E-14
	GO:0015672	1.44E-17	2.90E+01	1.49E+02	monovalent inorganic cation transport	BP	3.46E-14
	GO:0016021	3.05E-17	1.37E+02	2.32E+03	integral component of membrane	CC	6.41E-14
	GO:0099131	6.19E-17	1.40E+01	2.10E+01	ATP hydrolysis coupled ion transmembrane transport	BP	1.04E-13
	GO:0099132	6.19E-17	1.40E+01	2.10E+01	ATP hydrolysis coupled cation transmembrane transport	BP	1.04E-13
11	GO:0006811	3.80E-19	5.50E+01	5.34E+02	ion transport	BP	6.39E-15
	GO:0015988	2.09E-18	1.40E+01	1.90E+01	energy coupled proton transmembrane transport, against electrochemical gradient	BP	1.17E-14
	GO:0015991	2.09E-18	1.40E+01	1.90E+01	ATP hydrolysis coupled proton transport	BP	1.17E-14
	GO:0090662	4.57E-18	1.50E+01	2.40E+01	ATP hydrolysis coupled transmembrane transport	BP	1.92E-14
	GO:0099131	1.34E-17	1.40E+01	2.10E+01	ATP hydrolysis coupled ion transmembrane transport	BP	3.75E-14
	GO:0099132	1.34E-17	1.40E+01	2.10E+01	ATP hydrolysis coupled cation transmembrane transport	BP	3.75E-14
	GO:0015672	9.40E-17	2.70E+01	1.49E+02	monovalent inorganic cation transport	BP	2.26E-13
	GO:0005576	7.58E-16	5.70E+01	7.41E+02	extracellular region	CC	1.59E-12
	GO:0005215	1.18E-15	6.60E+01	8.91E+02	transporter activity	MF	2.20E-12
	GO:0044425	2.92E-15	1.48E+02	3.12E+03	membrane part	CC	4.91E-12
12	GO:0006811	1.60E-16	5.10E+01	5.34E+02	ion transport	BP	2.69E-12

	GO:0031224	7.01E-15	1.23E+02	2.40E+03	intrinsic component of membrane	CC	5.47E-11
	GO:0022857	9.76E-15	5.50E+01	6.93E+02	transmembrane transporter activity	MF	5.47E-11
	GO:0016021	3.38E-14	1.19E+02	2.32E+03	integral component of membrane	CC	1.27E-10
	GO:0005215	3.79E-14	6.30E+01	8.91E+02	transporter activity	MF	1.27E-10
	GO:0044425	5.41E-14	1.44E+02	3.12E+03	membrane part	CC	1.52E-10
	GO:0022804	1.41E-13	3.40E+01	3.02E+02	active transmembrane transporter activity	MF	3.38E-10
	GO:0005576	4.28E-13	5.20E+01	7.41E+02	extracellular region	CC	8.99E-10
	GO:0034220	1.01E-12	2.80E+01	2.21E+02	ion transmembrane transport	BP	1.71E-09
	GO:0022891	1.02E-12	4.70E+01	5.94E+02	NA	NA	1.71E-09
13	GO:0031224	3.60E-19	1.48E+02	2.40E+03	intrinsic component of membrane	CC	5.36E-15
	GO:0044425	6.38E-19	1.75E+02	3.12E+03	membrane part	CC	5.36E-15
	GO:0016021	3.32E-18	1.43E+02	2.32E+03	integral component of membrane	CC	1.86E-14
	GO:0015988	1.51E-17	1.40E+01	1.90E+01	energy coupled proton transmembrane transport, against electrochemical gradient	BP	5.08E-14
	GO:0015991	1.51E-17	1.40E+01	1.90E+01	ATP hydrolysis coupled proton transport	BP	5.08E-14
	GO:0090662	3.87E-17	1.50E+01	2.40E+01	ATP hydrolysis coupled transmembrane transport	BP	1.08E-13
	GO:0099131	9.86E-17	1.40E+01	2.10E+01	ATP hydrolysis coupled ion transmembrane transport	BP	2.07E-13
	GO:0099132	9.86E-17	1.40E+01	2.10E+01	ATP hydrolysis coupled cation transmembrane transport	BP	2.07E-13
	GO:0006811	1.03E-15	5.40E+01	5.34E+02	ion transport	BP	1.92E-12
	GO:0015672	2.92E-14	2.60E+01	1.49E+02	monovalent inorganic cation transport	BP	4.90E-11

[a] BP, biological process; CC, cellular component; MF, molecular function

Supplementary Table 10. Genes for core transporters expressed in the pleuropodia.

Protein	Number of transcripts	Upregulated	Downregulated	*Drosophila* putative homologs of upreguleted transcripts[a]
V-ATPase	17	15	0	(subunits) Vha68-2, Vha55, VhaSFD, Vha44, Vha36-1, Vha26, Vha14-1, Vha13, Vha100-1, Vha100-2, Vha16-1, VhaPPA1-1, VhaM9.7-a, VhaAC39-1, VhaAC45
Na+,K+ATPase	4	3	0	(subunit alpha) Atpalpha, (subunits beta) nrv2, nrv3
Carbonic anhydrase	5	3	0	CAH1, CAH2, CG3940
Cation/proton antiporters	8	2	0	Nhe3, Nha1
Other exchangers and cotransporters	14	2	3	Prestin, NKCC
Water channels	11	2	2	Eglp2, Eglp3
Potassium channels	22	1	1	Ork1
Chloride channels	9	0	0	

[a] Uniprot top blast hit

Supplementary Table 11. *Schistocerca* gene for proteins with GO "proteolysis" that were upregulated in the highly secreting pleuropodia.

***Schistocerca gregaria* transcript**		***Tribolium castaneum* top blast hit**			***Drosophila melanogaster* topblast hit**		
Transcript ID	**Protein**	**Uniprot ID**	**e-value**	**Protein**	**Uniprot ID**	**e-value**	**Protein**
SgreTb00181983	Aminopeptidase	D6WCY0	0	Aminopeptidase	Q7KRW4	0	Aminopeptidase (EC 3.4.11.-)
SgreTa00069 80	Aminopeptidase	D7EJC6	0	Aminopeptidase	Q86NQ5	0	Aminopeptidase (EC 3.4.11.-)
SgreTa00140 09	Angiotensin-converting enzyme	D6X4L0	0	Angiotensin-converting enzyme	X2J8C3	0	Angiotensin-converting enzyme (EC 3.4.-.-)
SgreTa00002 84	ATP-dependent Clp protease ATP-binding subunit clpX	A0A139WNK6	0	ATP-dependent Clp protease ATP-binding subunit clpX-like, mitochondrial	Q960M5	9.24E-175	LD45279p
SgreTa00167 82	Carbohydrate sulfotransferase	D6WTL6	1.20E-48	Carbohydrate sulfotransferase	Q9W070	2.20E-40	Carbohydrate sulfotransferase (EC 2.8.2.-)
SgreTc00000 11	Carboxypeptidase	D2A5G0	1.65E-19	Carboxypeptidase A	P42787	0.24	Carboxypeptidase D (EC 3.4.17.22) (Metallocarboxypeptidase D) (Protein silver)
SgreTa00144 01	Cathepsin B	D6WGZ1	4.52E-152	Cathepsin B	Q9VY87	6.74E-152	Cathepsin B1, isoform A (Cathepsin B1, isoform C) (EC 3.4.-.-) (EC 3.4.22.-) (GH06546p)
SgreTd0014 041	Cysteine proteinase	D6WPZ3	1.6	Cystatin	Q9VN93	1.93E-118	Putative cysteine proteinase CG12163 (EC 3.4.22.-)
SgreTa00026 70	E3 ubiquitin-protein ligase	D6WQG3	7.43E-13	E3 ubiquitin-protein ligase HRD1-like Protein	A0A0B4KHH2	1.96E-06	Septin interacting protein 3, isoform B (EC 6.3.2.-)
SgreTa00071 52	E3 ubiquitin-protein ligase synoviolin b-like	D6WQG3	0	E3 ubiquitin-protein ligase HRD1-like Protein	A0A0B4KHH2	0	Septin interacting protein 3, isoform B (EC 6.3.2.-)
SgreTa00046 53	Heat shock protein 90	D6X0J9	0	Heat shock protein 83-like Protein	Q9VAY2	0	Glycoprotein 93 (LD23641p)
SgreTa00025 96	Lys-63-specific deubiquitinase	D6X1A0	1.46E-43	Lys-63-specific deubiquitinase BRCC36-like Protein	Q9V3H2	5.38E-06	26S proteasome non-ATPase regulatory subunit 14 (EC 3.4.19.-) (26S proteasome regulatory complex subunit p37B) (26S proteasome regulatory subunit rpn11) (Yippee-interacting
SgreTa00176 92	Neprilysin	A0A139WI73	0	Neprilysin-2-like Protein	A0A0B4K692	0	Neprilysin-2 (EC 3.4.24.11) [Cleaved into: Neprilysin-2, soluble form]

SgreTa0017692	Neprilysin	A0A139WI73	0	Neprilysin-2-like Protein	A0A0B4K692	0	Neprilysin-2 (EC 3.4.24.11) [Cleaved into: Neprilysin-2, soluble form]
SgreTa0017746	Neprilysin	A0A139WHP0	1.37E-54	Neprilysin-2-like Protein	A0A0B4K692	1.38E-52	Neprilysin-2 (EC 3.4.24.11) [Cleaved into: Neprilysin-2, soluble form]
SgreTa0002467	Neprilysin	A0A139WHP0	0	Neprilysin-2-like Protein	A0A0B4K692	0	Neprilysin-2 (EC 3.4.24.11) [Cleaved into: Neprilysin-2, soluble form]
SgreTb0039045	Neprilysin	A0A139WI73	6.32E-14	Neprilysin-2-like Protein	A0A0B4K692	1.08E-12	Neprilysin-2 (EC 3.4.24.11) [Cleaved into: Neprilysin-2, soluble form]
SgreTb0039123	Neprilysin	A0A139WHP0	3.79E-23	Neprilysin-2-like Protein	A0A0B4K692	3.90E-26	Neprilysin-2 (EC 3.4.24.11) [Cleaved into: Neprilysin-2, soluble form]
SgreTa0016582	Protein roadkill-like	A0A139WMR0	2.72E-12	Protein roadkill-like Protein	C7LAF6	1.02E-15	RE09961p
SgreTa0017657	Putative serine protease, K12H4.7-like	D6WGL2	2.66E-101	Putative serine protease K12H4.7-like Protein	Q9VS02	5.88E-90	CG9953
SgreTa0004939	Rhomboid-like protein	D6WUJ2	1.61E-105	Rhomboid-like protein	Q9VYW6	1.39E-86	Rhomboid-like protein (EC 3.4.21.-)
SgreTb0023745	RING finger domain protein	D2A2S1	3.52E-13	RING-box protein 1A-like Protein	Q7JWH5	2.50E-32	MIP07211p (RE61847p) (Regulator of cullins 2, isoform A) (Regulator of cullins 2, isoform B)
SgreTa0005905	Selenoprotein S	D6WGX2	2.27E-06	Uncharacterized protein	Q9W0D3	6.6	GH15728p
SgreTa0017649	Serine protease	A0A139W9L3	1.55E-74	Serine protease P43	Q9VAQ3	5.03E-52	GH18608p
SgreTa0010219	Serine protease, Easter-like	D6WUF6	1.90E-20	Serine protease P136	P13582	4.44E-18	Serine_protease_easter_(EC_3.4.21.-)
SgreTb0037249	Serine protease, Easter-like	D6WUF7	9.05E-54	Serine protease H137	A0A0B4KGQ4	4.24E-52	Easter, isoform B (EC 3.4.-.-) (EC 3.4.21.-)
SgreTb0039024	Serine protease, Easter-like	D6WGT8	2.22E-26	Serine protease H33	P13582	1.35E-21	Serine protease easter (EC 3.4.21.-)
SgreTb0039879	Serine protease, Easter-like	D6WP87	2.25E-16	Serine protease P90	P13582	7.93E-14	Serine protease easter (EC 3.4.21.-)
SgreTa0003661	Serine protease, Easter-like	D6WUF6	5.95E-90	Serine protease P136	P13582	2.76E-82	Serine protease easter (EC 3.4.21.-)
SgreTa0007424	Serine protease, Easter-like	D6WUF6	2.87E-75	Serine protease P136	P13582	3.76E-76	Serine protease easter (EC 3.4.21.-)
SgreTa0007425	Serine protease, Easter-like	D6WUF6	1.68E-73	Serine protease P136	P13582	2.34E-75	Serine protease easter (EC 3.4.21.-)
SgreTa0007426	Serine protease, Easter-like	D6WUF6	3.37E-60	Serine protease P136	A0A126GUP6	5.60E-61	Melanization protease 1 (EC 3.4.21.-)
SgreTa0001065	Serine protease, gd-like	D6WYU8	3.73E-56	Serine protease P69	A4V9W2	2.70E-37	CG9649_protein
SgreTb0003860	Serine protease, H2-like	D6WBT0	1.02E-81	Serine protease H2	Q86PE8	1.14E-69	SD23103p

Supplementary Table 12. Differential gene expression of *Schistocerca* homologs (or close relatives) of known genes for MF proteases in the intensively secreting pleuropodia.

MF protein[a]	Blast query NCBI, SilkDB ID[b]	Blast query Uniprot ID	*Schisctocerca* transcript ID	RPKM leg	RPKM plp	prob[c]	UP[d]	DOWN[e]	*Bombyx mori* Uniprot ID	e-value	*D. melanogaster* Uniprot ID	e-value
Putative peptidase	ADA67926.1	D2KMR2	SgreTa0000627	41.926	131.748	0.849	3.142	0.318	H9JHZ1	0	Q9V3U6	0
Molting fluid carboxypeptidase A precursor	NP_001036933.1/BGIBMGA008910	Q60F93	SgreTa0007896	34.902	0.075	0.944	0.002	466.633	Q60F93	5.65E-131	Q9W475	3.25E-121
			SgreTa0004360	1.343	0.181	0.426	0.135	7.430	Q60F93	7.56E-82	Q961J8	3.19E-92
			SgreTa0005012	filtered out					Q60F93	6.85E-62	Q961J8	1.04E-76
			SgreTa0010262	filtered out					Q60F93	3.95E-66	Q961J8	7.84E-78
			SgreTa0012118	0.404	0.191	0.144	0.474	2.112	Q60F93	1.26E-51	Q9W475	2.53E-48
Carboxypeptidase	NP_001091798.1	A0FDQ4	SgreTa0011942	filtered out					H9BVM4	4.17E-101	Q9VL86	5.43E-111
Aminopeptidase N-12	AFK85028.1	I3VR83	SgreTa0000698	2.333	1.673	0.235	0.717	1.395	I3VR83	0	Q0KI00	0
			SgreTb0018983	8.251	35.863	0.854	4.347	0.230	I3VR83	0	Q7KRW4	0
			SgreTa0006248	11.151	0.111	0.848	0.010	100.162	I3VR83	1.17E-58	Q8IGR1	3.38E-64
Neutral endopeptidase 24.11	NP_001036959.1	Q9BLH1	SgreTa0002467	62.256	2282.007	0.987	36.655	0.027	H9ITE9	0	A0A0B4K692	0
			SgreTa0017692	0.555	133.297	0.982	240.283	0.004	H9ITE9	3.8	A0A0B4K692	0
			SgreTa0014732	5.410	0.010	0.745	0.002	546.280	H9ITE9	0	Q9W436	0
			SgreTa0007268	4.191	0.194	0.688	0.046	21.661	H9ITE9	1.72E-179	Q9W436	3.55E-169

Aminopeptidase N precursor	NP_001037013.1	O76803	SgreTa0004368	5.077	12.821	0.687	2.525	0.396	Q9BLH1	6.30E-23	Q9W5Y0	5.44E-27
			SgreTb0039123	1.173	219.355	0.989	186.958	0.005	Q9BLH1	4.05E-27	A0A0B4K692	3.90E-26
Ecdysteroid-inducible angiotensin-converting enzyme-related gene product precursor	NP_001036859.1	Q9NDS8	SgreTa0014009	65.757	1457.474	0.979	22.165	0.045	H9IZ41	0	X2J8C3	0
			SgreTa0017728	1.099	62.706	0.962	57.083	0.018	H9IZ41	0	Q8SXX2	0
prolyl endopeptidase/PREDICTED: prolyl endopeptidase	BGIBMGA002593-PA/XP_012551091.1	H9IZA8	SgreTa0014314	30.625	16.175	0.669	0.528	1.893	H9IZA8	0	Q9VKW5	0
Metalloendopeptidase /PREDICTED: zinc metalloproteinase nas-7-like	BGIBMGA013174-PA/XP_012552528.1	H9JUG1	SgreTa0008591	11.615	0.053	0.854	0.005	220.844	H9JUG1	6.98E-44	Q9VC98	1.51E-83
Peptidase_M14 / Carboxypeptidase M	BGIBMGA005320	H9J728	SgreTa0000995	64.330	10.542	0.905	0.164	6.102	H9J728	5.66E-158	B7Z0Z5	0
			SgreTa0016996	7.454	18.780	0.724	2.519	0.397	H9J728	4.28E-99	A0A023GPN7	9.36E-126
Peptidase_M14 / Carboxypeptidase D-like	BGIBMGA012806	H9JTE4	SgreTc0000011	23.158	72.928	0.838	3.149	0.318	H9JTE5	0	P42787	0

[f] nf

Peptidase_M14 / Carboxypeptidase-like	BGIBMGA006715	H9JB20	SgreTa0000925	nf	12.767	139.807	0.948	10.950	0.091	H9IST0	3.60E-167	D3DME3	8.86E-87
Peptidase_M14 / Carboxypeptidase E-like	BGIBMGA000307	H9IST0	SgreTa0000925		12.767	139.807	0.948	10.950	0.091	H9IST0	3.60E-167	D3DME3	8.86E-87
Peptidase_M14 / Venom serine carboxypeptidase-like	BGIBMGA003112	H9JOS5	SgreTa0007812		48.723	80.870	0.640	1.660	0.602	H9JOS5	0	Q9VIK1	0
Peptidase_M14 / Carboxypeptidase B	BGIBMGA001891	H9IXA8	SgreTa0005501		filtered out					H9IXA8	0.14	Q9VNR6	6.3
Peptidase_M14 / Angiotensin-converting enzyme-like	BGIBMGA002526	H9IZ41	SgreTa0003298		4.184	23.637	0.854	5.649	0.177	H9IZ41	4.80E-24	M9PCE8	8.60E-23
Peptidase_M1 / Aminopeptidase N-like	BGIBMGA008066	H9JEW9	SgreTa0017219		0.893	391.034	0.994	437.926	0.002	H9JEW9	5.02E-27	Q9VAP0	1.98E-132
Peptidase_M1 / Aminopeptidase N	BGIBMGA010764	H9JML1		nf									
Peptidase_C1 / Peptidase C1 (cathepsin B)	BGIBMGA004421	H9J4I1		nf									
Peptidase_C1 / Digestive cysteine protease 1 (cathepsin L)	BGIBMGA009139	H9JHZ1	SgreTa0000627		41.926	131.748	0.849	3.142	0.318	H9JHZ1	0	Q9V3U6	0

Peptidase_C1 / Digestive cysteine protease 1 (cathepsin L)	BGIBMGA009184	D2KMR2	nf												5.80E-28
Peptidase_C1 / cathepsin B	BGIBMGA007061	H9JC16	SgreTa0002023	filtered out									Q9VY87		
Peptidase_S28 / Serine carboxypeptidase	BGIBMGA003579	H9J242	SgreTa0017657	2.309	904.260	0.996	391.599	0.003					Q9VS02	4.69E-77	5.88E-90
Serine protease-like protein precursor	NP_001037368.1	Q45RG0	SgreTa0009552	11.605	0.022	0.855	0.002	532.339					Q9VUG2	7.21E-26	1.23E-145
			SgreTa0000099	25.653	0.164	0.926	0.006	156.171					Q9I7V4	2.16E-33	3.96E-32
			SgreTa0006919	13.052	0.218	0.865	0.017	59.881					E1JIW3	1.83E-68	2.62E-110
			SgreTa0007301	filtered out									Q8IQ10	7.68E-42	2.70E-46
			SgreTb0018716	filtered out									Q9VHF7	3.31E-10	2.46E-15
Masquerade-like serine proteinase homolog, Serine protease homolog 1	NP_001037053.1	Q8I9N4	SgreTa0003265	44.750	24.552	0.672	0.549	1.823					Q8I9N4	8.75E-130	2.34E-122
			SgreTa0000788	12.832	0.294	0.861	0.023	43.619					Q9VL01	3.43E-16	7.47E-138
Serine protease	NP_001036903.1	Q762I9	nf										Q9VV38	H9JQC6	
Scolexin	NP_001093269.1	Q1HPY5	nf												
serine protease HP21	BGIBMGA009610-PA/XP_012545963.1	H9JJA9	SgreTa0017649	7.350	179.691	0.972	24.446	0.041					F5BZV1	4.84E-66	5.03E-52
			SgreTa0001854	0.500	1.238	0.307	2.476	0.404					H9JJA9	1.34E-08	8.36E-09
			SgreTb0011715	0.177	1.029	0.360	5.803	0.172					H9JJA9	2.32E-07	7.54E-05

Chymotrypsin-like proteinase	AFD99127.1	H9BVM5	nf				0.996	153.480	0.007								
30kP protease A (43k peptide) precursor/PREDICTED: serine protease 3-like	BGIBMGA012787-PA/XP_004928982.1	H9JTC5	nf														
Cocoonase	BAJ46146.1	E5RVK3	nf														
35kDa protease precursor	NP_001037037.1	Q9BMQ7	SgreTa0011958	filtered out						Q9BMQ7	2.39E-23	Q9VRS6	2.35E-30				
Trypsin-like protease precursor	NP_001093273.1	Q1HPT9	nf														
BGIBMGA011362-PA (trypsin-like serine protease - fibroin heavy chain) /PREDICTED: uncharacterized protein LOC101736544 [Bombyx mori]	BGIBMGA011362-PA/XP_004930740.1	H9JPA8	SgreTa0001636	49.377	7578.311					H9JPA8	1.44E-08	Q95RS6	1.20E-22				
BGIBMGA000840-PA (trypsin-like protease)/PREDICTED: transmembrane protease serine 9-like	BGIBMGA000840-PA/XP_012547171.1	H9IUB0	SgreTa0011597	filtered out						H9IUB0	2.03E-21	A4V9Z3	3.53E-24				
			SgreTa0012375	filtered out						H9IUB0	8.80E-16	Q9VEA0	4.28E-19				
Proclotting enzyme-like[g]	BGIBMGA003461	H9J1S4	SgreTc0000001	0.247	0.500	0.157	2.023	0.494		H9J1S4	4.10E-37	Q8T8X4	5.16E-103				
			SgreTa0015939	0.351	0.500	0.102	1.426	0.701		H9J1S4	5.63E-19	Q9VK10	0.052				

BGIBMGA013738-PA (hemolymph proteinase 16 [Manduca sexta] - coagulation factor X)/PREDICTED: transmembrane protease serine 11A-like	BGIBMGA013738-PA/XP_012553183	H9JW25	nf															1.54E-144	Q8SY35
BGIBMGA000841-PA (trypsin-2)/PREDICTED: transmembrane protease serine 9-like [Bombyx mori]	BGIBMGA000841-PA/XP_012547171.1	H9IUB1	nf															2.59E-19	A0A0B4LFV3
clip-SP / Serine protease Stubble	BGIBMGA009743	H9JJP2		SgreTa0002070	70.633	0.079	0.970	0.001	892.952	4.40E-14	H9JJP2	7.18E-19	H9JJP2						
clip-SP / Serine protease, Snake-like	BGIBMGA010306	H9JLA5	nf	SgreTa0006098	23.607	0.358	0.918	0.015	65.882										
clip-SP / Serine protease, Easter-like	BGIBMGA013746	Q9XXV0	nf																
clip-SP / Clip domain serine protease 3	BGIBMGA010257	H9JL56		SgreTa0011544	1.137	0.186	0.382	0.163	6.117	1.66E-93	H9JL56	2.16E-88	Q9VW19						
clip-SP / Serine protease, Easter-like	BGIBMGA005173	Q2VG86		SgreTa0003188	0.580	485.973	0.995	837.454	0.001	1.39E-66	Q2VG86	4.97E-61	A0A126GUP6						
				SgreTa0003661	29.498	1332.786	0.989	45.183	0.022	1.08E-77	Q2VG86	2.76E-82	P13582						
				SgreTa0006780	7.003	103.368	0.953	14.762	0.068	5.36E-57	Q2VG86	2.50E-63	A0A126GUP6						

clip-SP / Serine protease, Easter-like	BGIBMGA010546	H9JLZ4	SgreTa0007424	0.374	29.615	0.933	79.133	0.013	Q2VG86	1.15E-73	P13582	3.76E-76
			SgreTa0007425	1.711	123.690	0.978	72.307	0.014	Q2VG86	4.20E-75	P13582	2.34E-75
			SgreTb0037249	0.087	21.762	0.915	249.740	0.004	Q2VG86	4.51E-51	A0A0B4KGQ4	4.24E-52
			SgreTb0039879	0.562	305.632	0.993	544.040	0.002	Q2VG86	3.75E-11	P13582	7.93E-14
			SgreTa0007304	0.140	5.787	0.751	41.323	0.024	H9JLZ4	4.36E-09	A0A126GUP6	5.90E-09
			SgreTa0010219	2.223	46.116	0.941	20.746	0.048	H9JLZ4	3.12E-17	P13582	4.44E-18
			SgreTb0039024	0.529	11.706	0.843	22.114	0.045	H9JLZ4	7.32E-22	P13582	1.35E-21
clip-SPH / Serine protease 42-like	BGIBMGA009551	Q1HPQ5	nf									
clip-SPH / Serine protease Stubble	BGIBMGA011732	H9JQC6	SgreTa0000788	12.832	0.294	0.861	0.023	43.619	H9JQC6	3.43E-16	Q9VV38	7.47E-138
clip-SPH / Serine protease 1	BGIBMGA014404	H9JXY6	SgreTb0003860	77.416	1727.410	0.979	22.313	0.045	H9JXY6	0.097	Q86PE8	1.14E-69
clip-SPH / Scarface, partial	BGIBMGA011363	H9JPA9	nf									
GD_N-SP / Chymotrypsin-C	BGIBMGA013049	H9JU37	SgreTa0007959	filtered out					H9JU37	5.49E-07	Q24111	3.13E-11
			SgreTa0014461	11.867	0.013	0.858	0.001	928.634	H9JU37	2.58E-72	O62589	1.91E-47
GD_N-SP / Serine protease gd	BGIBMGA013051	H9JU39	nf									
LDLa-CCP-SP / Pattern recognition serine protease precursor	BGIBMGA012217	H9JRQ7	nf									

SP / Serine protease Stubble	BGIBMGA009750	H9JJP9	SgreTa0005037	21.941	0.701	0.907	0.032	31.317	H9JJP9	3.68E-10	A1Z7M3	5.07E-12		
			SgreTa0005935	0.338	2.199	0.518	6.513	0.154	H9JJP9	7.52E-56	A0A0B4LH73	5.03E-60		
			SgreTa0007005	31.493	0.551	0.935	0.018	57.109	H9JJP9	4.11E-141	A1Z7M3	4.21E-151		
SP / Serine protease, Snake-like	BGIBMGA012427	H9JSB6	nf						H9IWW2	1.09E-83	A8JQZ2	1.61E-72		
SP / Transmembrane protease serine 9-like	BGIBMGA006272	H9J9S9	nf					1.728	H9IWW2	6.72E-89	A8JQZ2	5.48E-69		
SP / Serine protease, Snake-like	BGIBMGA001745	H9IWW2	SgreTa0000782	0.347	0.201	0.109	0.579	0.005	H9IWW2	4.06E-10	Q9VMZ3	8.01E-17		
			SgreTa0000783	4.295	917.473	0.995	213.590	0.308	H9IWW2	1.20E-41	Q9VWU1	8.75E-45		
			SgreTa0001853	0.500	1.625	0.394	3.250	2.268	H9IWW2	9.42E-64	A8JQZ2	1.30E-55		
			SgreTa0004364	7.715	3.402	0.596	0.441	0.273						
			SgreTa0008277	3.585	13.122	0.754	3.661							
SP / Serine protease, Snake-like	BGIBMGA012423	H9JSB2	nf											
SP / Trypsin 3A1-like	BGIBMGA000841	H9IUB1	nf											
SP / Chymotrypsin-1-like	BGIBMGA013962	H9JWP8	nf											
SPH / Transmembrane protease serine 9-like	BGIBMGA006406	H9JA61	SgreTa0009552	11.605	0.022	0.855	0.002	532.339	H9JA61	7.21E-26	Q9VUG2	1.23E-145		
SPH / Scolexin B	BGIBMGA006423	Q1HPY5	nf											

SPH / Serine protease Snake	BGIBMGA011051	H9JNE8	nf								2.54E-18
SPH / Chymotrypsin-like elastase family member 2A	BGIBMGA013738		nf							Q7K3Y1	
		H9JW25									
SPH / Serine protease 1	BGIBMGA014407	H9JXY9	nf								
SPH / Serine protease, Easter-like	BGIBMGA005172	H9J6N0	nf								
SPH / Serine protease, Easter-like	BGIBMGA013797	H9JW83	nf								
SPH / Chymotrypsin-like elastase family member 2A	BGIBMGA013738	H9JW25	nf								
SPH / Serine protease, Snake-like	BGIBMGA012425	H9JSB4	nf								
SPH / Trypsin-1-like	BGIBMGA006747	H9JB52	nf								
SPH / Hemolymph protease 6	BGIBMGA005380	H9J788	SgreTb0019518	2.758	0.945	0.478	0.343	2.919	H9J788	1.05E-22	

[a] sequences for search from the MF of the lepidopteran *Bombyx mori* (Zhang et al., 2014; Liu et al., 2018)
[b] from original publications
[c] values lower than threshold in grey
[d] significant upregulation in magenta
[e] significant downregulation in blue
[f] not found

Supplementary Table 13. Immunity-related proteins expressed in the highly secreting pleuropodia.

Transcript ID[a]	Protein	Characteristics	RPKM LEG	RPKM PLP	prob	Fold UP	*D. melanogaster* Uniprot ID	e-value	*Locusta migratoria* Uniprot ID	e-value
SgreTa0009559	Gram-negative bacteria binding protein 3	recognition	15.396	510.045	0.984	33.128	A8E788	4.61E-72	H8YU83	0
SgreTa0001756	Gram-negative bacteria-binding protein 3	recognition	9.729	33.740	0.824	3.468	A8E788	6.59E-25	H8YU81	1.31E-51
SgreTa0008497	Peptidoglycan-recognition protein, SA-like	recognition	4.348	181.319	0.980	41.704	X2JEI8	7.49E-64	H8YU85	1.56E-115
SgreTa0017598	Peptidoglycan-recognition protein	recognition	2.891	20.032	0.855	6.930	Q95T64	1.3	A0A1J0M093	3.75E-55
SgreTa0017671	Peptidoglycan-recognition protein LE	recognition	1.817	18.691	0.869	10.286	M9NDY2	0.66	A0A1S6M249	3.36E-160
SgreTa0008286	Peptidoglycan-recognition protein LC	recognition	2.068	17.910	0.858	8.660	E1JI88	4.01E-32	A0A1S6M249	0.009
SgreTa0016070	C-type lectin	recognition	0.256	210.951	0.990	824.507	Q9VGA3	7.69E-61	Q6RX62	1.2
SgreTa0004881	C-type lectin	recognition	0.299	185.308	0.988	619.984	Q9VGA3	9.19E-61	Q6RX62	0.26
SgreTa0008846	C-type lectin	recognition	0.101	73.255	0.971	722.715	Q9VQ53	1.77E-09	A0A0M3SBM6	0.79
SgreTa0017398	C-type lectin	recognition	0.602	54.654	0.960	90.833	Q9VQ53	1.16E-08	W5U4R1	3.5
SgreTa0017399	C-type lectin	recognition	0.524	40.039	0.948	76.477	Q9W0I2	8.78E-06	W5U4R1	3.5
SgreTa0017779	C-type lectin	recognition	0.087	35.062	0.944	404.595	Q9VLW1	1.33E-09	X5MBK1	1.6
SgreTa0007672	C-type lectin	recognition	0.136	34.664	0.943	255.779	B7YZU7	0.009	I3RT27	0.64
SgreTa0007333	C-type lectin	recognition	3.176	30.511	0.899	9.608	Q9VKL8	2.03E-08	V5RDW2	3.8
SgreTa0013807	Sialomucin	recognition - adhesion	87.182	242.138	0.832	2.777	Q9VT37	1.07E-14	A0A0M4J319	2.3
SgreTa0015095	Cactus	Toll signaling	36.608	142.408	0.879	3.890	C1C5B1	5.12E-64	L7WUW9	0.35
SgreTa0017744	Spaetzle	serine protease - Toll signaling	0.793	14.559	0.863	18.362	Q8IMP8	2.90E-17	L7WRS0	7.3
SgreTa0007426	Serine protease, Easter-like	serine protease - Toll signaling	0.658	673.428	0.996	1023.601	A0A126GUP6	5.60E-61	A8QL65	4.12E-30
SgreTa0003661	Serine protease, Easter-like	serine protease - Toll signaling	29.498	1332.786	0.989	45.183	P13582	2.76E-82	A8QL65	8.32E-42

SgreTb0039879	Serine protease, Easter-like	serine protease - Toll signaling	0.562	305.632	0.993	544.040	P13582	7.93E-14	A8QL65	3.62E-05
SgreTa0007425	Serine protease, Easter-like	serine protease - Toll signaling	1.711	123.690	0.978	72.307	P13582	2.34E-75	A8QL65	2.01E-37
SgreTa0010219	Serine protease, Esater-like	serine protease - Toll signaling	2.223	46.116	0.941	20.746	P13582	4.44E-18	X5MNU5	5.27E-09
SgreTa0007424	Serine protease, Esater-like	serine protease - Toll signaling	0.374	29.615	0.933	79.133	P13582	3.76E-76	A8QL65	3.58E-36
SgreTb0037249	Serine protease, Easter-like	serine protease - Toll signaling	0.087	21.762	0.915	249.740	A0A0B4KGQ4	4.24E-52	A8QL65	1.57E-24
SgreTb0039024	Serine protease, Easter-like	serine protease - Toll signaling	0.529	11.706	0.843	22.114	P13582	1.35E-21	A8QL65	1.83E-08
SgreTa0017715	Serine protease, Snake-like	serine protease - Toll signaling	70.550	2947.456	0.989	41.778	A8JQZ2	1.97E-31	A8QL65	2.94E-15
SgreTa0000783	Serine protease, Snake-like	serine protease - Toll signaling	4.295	917.473	0.995	213.590	A8JQZ2	2.5	A8QL65	7.97E-39
SgreTa0017635	Serine protease, Snake-like	serine protease - Toll signaling	0.687	321.676	0.993	468.177	A8JQZ2	2.26E-39	X5MI45	8.67E-21
SgreTa0017649	Serine protease HP21	serine protease - Toll signaling	7.350	179.691	0.972	24.446	Q9VAQ3	5.03E-52	A8QL65	1.45E-35
SgreTa0001065	Serine protease	serine protease - Toll signaling	0.060	32.068	0.940	530.802	A4V9W2	2.70E-37	A8QL65	1.15E-20
SgreTa0009747	Serpin (27-like)	response, serpin - melanization, Toll signaling	14.486	1054.669	0.992	72.804	Q9V3N1	7.46E-63	L7WRS0	1.06E-54
SgreTa0006966	Pro-phenol oxidase subunit 2	response - melanization	144.781	1347.432	0.954	9.307	Q9V521	0	C0LV92	0
SgreTa0003188	Serine protease (Melanization protease), hemolymph proteinase 8-like	response - melanization, serine protease, Toll signaling?	0.580	485.973	0.995	837.454	A0A126GUP6	4.97E-61	A8QL65	5.74E-27
SgreTa0006780	Serine protease (Melanization protease 1)	response - melanization, serine protease	7.003	103.368	0.953	14.762	A0A126GUP6	2.50E-63	A8QL65	3.30E-37

SgreTa0003769	Serine protease (Melanization protease 1)	response - melanization, serine protease	14.916	57.352	0.859	3.845	A0A126GUP6	2.50E-66	A8QL65	2.45E-38
SgreTa0015666	Serine protease (Melanization protease 1)	response - melanization, serine protease	0.311	17.443	0.894	56.048	A0A126GUP6	8.34E-27	X5MNU3	6.00E-19
SgreTa0015520	Protein yellow	response - melanization	2.751	520.094	0.993	189.076	Q9V4C0	7.45E-115	A0A0M4JNZ6	1.5
SgreTa0001721	Atypical protein kinase C	response - melanization	29.355	199.268	0.931	6.788	A1Z9X0	0	D6BL32	3.31E-35
SgreTb0009421	Antimicrobial peptide, similar to diptericin	response - AMP	9.598	84.327	0.932	8.786	A7LFL7	2.4	K4PXJ9	1.57E-35
SgreTa0007897	C-type lysozyme	response - lysozyme	42.930	14452.147	0.998	336.642	Q4QPT0	7.19E-40	W0C415	7.68E-81
SgreTa0008528	C-type lysozyme	response - lysozyme	8.608	1159.550	0.995	134.708	Q4QPT0	4.89E-29	W0C415	1.72E-29
SgreTa0017707	I-type lysozyme	response - lysozyme	12.198	6712.313	0.998	550.259	Q7JYZ0	1.12E-36	A0A1J0M172	1.19E-32
SgreTb0006243	I-type lysozyme	response - lysozyme	16.959	519.352	0.982	30.623	Q7JYZ0	7.19E-32	A0A1J0M172	3.23E-22
SgreTa0017736	I-type lysozyme	response - lysozyme	1.338	257.839	0.990	192.682	Q8SY67	5.00E-32	A0A1J0M172	1.44E-21
SgreTa0012102	Nitric oxide synthase	response - RNS production	0.720	35.662	0.940	49.552	Q27571	8.81E-85	A0A1L2EC44	5.34E-32
SgreTa0012465	Nitric oxide synthase	response - RNS production	0.551	31.976	0.936	58.021	Q27571	3.70E-119	A0A1L2EC44	4.40E-25
SgreTa0011538	Nitric oxide synthase	response - RNS production	0.550	29.827	0.932	54.222	Q27571	7.76E-107	A0A1L2EC44	0.12
SgreTa0002218	Nitric oxide synthase	response - RNS production	0.483	26.606	0.925	55.098	Q27571	2.52E-79	A0A0M4J774	0.85
SgreTb0026778	Nitric oxide synthase	response - RNS production	0.609	22.343	0.910	36.701	Q27571	1.88E-45	A0A0M3SBP5	0.46
SgreTa0009095	Catalase	response - quenching of ROS	355.147	1158.274	0.867	3.261	P17336	0	A0A1B3PEJ4	0
SgreTa0017216	Glutathione S-transferase delta	response - quenching of ROS	44.465	121.782	0.821	2.739	P20432	3.41E-54	V9Q4A2	1.44E-117
SgreTa0001382	Glutathione S-transferase delta	response - quenching of ROS	6.581	34.561	0.870	5.252	P20432	6.77E-44	V9Q3Y3	1.30E-100
SgreTa0005082	Glutathione S transferase sigma	response - quenching of ROS	30.327	474.773	0.970	15.655	A4UZL5	5.76E-46	V9Q3X5	2.19E-117
SgreTa0001135	Glutathione S-transferase sigma	response - quenching of ROS	48.832	304.240	0.929	6.230	A4UZL5	2.11E-65	F4YUJ6	6.59E-117

SgreTa0001229	Glutathione S-transferase sigma	response - quenching of ROS	10.716	178.991	0.964	16.703	A4UZL5	3.27E-66	F4YUJ2	2.74E-125
SgreTa0001228	Glutathione S-transferase sigma	response - quenching of ROS	2.523	61.965	0.952	24.560	A4UZL5	1.20E-66	V9Q487	3.45E-137
SgreTa0015725	Glutathione S-transferase sigma	response - quenching of ROS	3.831	58.263	0.941	15.209	A4UZL5	7.15E-66	F4YUJ4	1.02E-125
SgreTa0008740	Glutathione S-transferase theta	response - quenching of ROS	11.795	141.826	0.952	12.025	Q9VRA4	7.23E-55	F4YUJ7	1.02E-150
SgreTa0000588	Glutathione S-transferase theta	response - quenching of ROS	12.479	108.672	0.936	8.708	Q8SZE4	2.25E-56	V9Q331	4.15E-148
SgreTb0009705	Glutathione peroxidase	response - quenching of ROS	39.962	107.085	0.815	2.680	Q9VZQ8	1.97E-32	C0LV92	1.3
SgreTa0013400	Peroxiredoxin, 5-like	response - quenching of ROS	101.103	1034.150	0.957	10.229	Q960M4	7.63E-77	Q9U943	1.2
SgreTa0003898	Peroxiredoxin	response - quenching of ROS	5.621	236.221	0.982	42.028	A1Z893	2.11E-80	L7WSH9	4.75E-17
SgreTa0017700	Peroxidase	response - quenching of ROS	5.357	874.514	0.994	163.250	A0A0B4KHM9	0	A0A1L4A1S2	2.57E-126
SgreTa0015795	Peroxidase	response - quenching of ROS	30.639	84.306	0.816	2.752	E1JIN0	1.42E-129	A0A1L4A1S2	7.67E-104
SgreTa0017852	Peroxidase	response - quenching of ROS	0.417	49.355	0.958	118.307	A0A0B4KHM9	6.97E-24	A0A1L4A1S2	2.49E-60
SgreTa0000923	Limulus clotting factor C	serine protease - coagulation	9.481	170.887	0.966	18.025	Q9VER6	0.00028	O46131	0.27
SgreTa0004939	Rhomboid-like protein	serine protease	0.040	11.486	0.853	289.830	Q9VYW6	1.39E-86	A0A0M3SBM9	0.76
SgreTa0001074	Serine protease, Stubble-like	serine protease	10.846	96.190	0.935	8.869	Q8MS52	9.38E-171	A8QL65	1.41E-48
SgreTa0001636	Serine protease	serine protease	49.377	7578.311	0.996	153.480	Q95RS6	1.20E-22	A8QL65	6.77E-09
SgreTb0003860	Serine protease, H2-like	serine protease	77.416	1727.410	0.979	22.313	Q86PE8	1.14E-69	A8QL65	7.42E-39
SgreTa0017657	Putative serine protease, K12H4.7-like	serine protease	2.309	904.260	0.996	391.599	Q9VS02	5.88E-90	A0A0M4IUC5	0.095
SgreTa0003047	Serine protease	serine protease	0.181	112.822	0.981	622.806	A1Z7D2	3.66E-65	A8QL65	9.48E-32
SgreTa0006539	Serpin, 88E-like	serpin	32.423	2152.138	0.993	66.377	Q9VFC2	1.29E-83	L7WRS0	3.49E-36
SgreTa0002155	Uncharacterized serine protease inhibitor	serpin	33.831	646.731	0.975	19.117	Q8MST1	8.7	A0A0B5HB40	0.26

SgreTa0004379	Serine protease inhibitor serpin	serpin	14.060	439.136	0.982	31.233	A4V9T4	0.012	L7WRS0	3.48E-18
SgreTa0001866	CD109 antigen	serpin	52.607	237.109	0.901	4.507	Q9NFV6	0	A0A0F6PMG3	2.7
SgreTa0001991	Serine protease inhibitor (Serpin-6)	serpin	1.040	209.051	0.988	201.047	Q7JV69	1.37E-34	L7WRS0	3.43E-31
SgreTa0005256	Serpin	serpin	3.825	204.845	0.984	53.549	Q7JV69	1.24E-51	L7WRS0	2.03E-39
SgreTa0006999	Serpin	serpin	1.247	188.064	0.987	150.843	Q8MPN6	1.48E-59	L7WRS0	2.83E-94
SgreTa0014760	Serpin	serpin	40.836	186.485	0.900	4.567	Q8MPN6	1.42E-70	L7WRS0	5.83E-119
SgreTa0000537	Serpin	serpin	0.555	131.732	0.982	237.449	Q8MPN6	1.02E-82	L7WRS0	9.12E-141
SgreTa0009508	Serpin	serpin	0.215	43.774	0.954	203.176	Q8MM39	1.56E-78	L7WRS0	1.05E-117
SgreTa0013328	Ferritin	response - iron sequestration	238.095	651.309	0.834	2.735	Q7KRU8	1.37E-68	A0A0M3SBM6	0.95
SgreTa0009587	Transferrin	response - iron sequestration	4.569	104.869	0.963	22.954	A1ZAC0	3.81E-148	A0A0K2D699	2.2
SgreTa0007180	Transcription factor AP-1 (Jun-related antigen)	TF - JNK signaling	53.712	226.132	0.893	4.210	P18289	5.60E-41	V5QNC4	0
SgreTa0008467	X-box-binding protein 1	TF	73.041	276.111	0.883	3.780	Q5BI44	6.75E-17	V5QNC4	0.002
SgreTa0001608	Uncharacterized, contains BNIP3 domain	TF	43.705	112.464	0.807	2.573	Q9VPD6	2.69E-35	Q9U0T5	5.7
SgreTa0017586	Insulin-like receptor	receptor	0.367	71.810	0.970	195.593	G7H807	1.29E-102	D6BL32	8.25E-06
SgreTa0017722	Ionotropic receptor, 21a-like	receptor	0.214	21.032	0.911	98.183	Q9VPI2	6.50E-56	A0A0M3SBL8	0.003
SgreTa0005568	Vascular endothelial growth factor receptor	receptor	3.254	20.393	0.850	6.267	B6IDV6	0	D6BL32	8.26E-09
SgreTa0017768	Arylsulfatase B		2.477	392.230	0.992	158.371	Q8SZ72	5.49E-52	A0A0C4G3S7	0.17
SgreTa0016783	Glutaminyl-peptide cyclotransferase		58.374	179.467	0.849	3.074	Q0GT94	5.27E-105	A0A0M4IU86	2.3
SgreTa0001361	Kinase		9.374	151.459	0.961	16.157	Q8SXS8	1.24E-151	A0A0M4JNQ3	1.4
SgreTa0014204	Zinc finger protein, Tis11-like		35.701	131.876	0.872	3.694	B3DN56	1.04E-39	V9TLV5	7

SgreTa0015984	Beta-hexosaminidase *(Sg-hex)*	2.325	110.148	0.974	47.372	Q8IRB6	1.24E-65	L0APJ7	5.65E-65
SgreTa0002990	Cellular repressor of E1A-stimulated genes	26.255	71.205	0.810	2.712	Q9VEK7	1.45E-26	A0A0M3SBP7	0.39
SgreTa0004373	Mannosyl-oligosaccharide alpha-1,2-mannosidase	21.423	70.209	0.843	3.277	P53624	0	W8E8I1	3.3
SgreTa0014591	Phospholipid-transporting ATPase	6.196	47.189	0.909	7.615	B7YZF6	0	W5U4R1	0.000645
SgreTa0015546	Uncharacterized, contains Calcium-binding EGF domain	3.969	22.069	0.847	5.560	M9NCR4	9.4	O46131	1.4

[a] transcripts in bold were among the top 10% most highly "expressed" upregulated DEGs (Table 1)

Supplementary Table 14. Genes for lysozymes identified in the Schistocerca gregaria embryonic transcriptome.

Lysozyme type	*Schistocerca* gene	Transcript ID	*D. melanogaster* Uniprot ID	e-value	Transcript length (bp)	Protein length (aa)	N-terminus	C-terminus	Signal peptide (position)[a]
C-type lysozyme	*Sg-LyzC-1*	SgreTa0007897	CG11159	7.19E-40	655	141	yes	yes	yes (aa6-aa17)
	Sg-LyzC-2	SgreTa0008528	CG11159	4.89E-29	630	141	yes	yes	yes (aa6-aa17)
I-type lysozyme	*Sg-LyzI-1*	SgreTa0017707	CG6426	1.12E-36	1014	148	yes	yes	yes (aa4-aa15)
	Sg-LyzI-2	SgreTb0019973	CG6426	5.49E-56	792	168	yes	yes	yes (aa6-aa17)
	Sg-LyzI-3	SgreTb0006243	CG6426	7.19E-32	1565	167	yes	yes	yes (aa14-aa25)
	Sg-LyzI-4	SgreTa0016727	CG6435	1.45E-23	702	161	yes	yes	yes (aa5-aa16)
	Sg-LyzI-5	SgreTb0010420	CG6421	2.55E-27	1173	162	no	yes	yes (aa5-aa20)
Lysozyme-like	*Sg-Lyz-like-1*	SgreTa0017736	CG6429	5.00E-32	885	165	yes	yes	yes (aa11-aa23)
	Sg-Lyz-like-2	SgreTb0026237	CG8492	2.95E-35	271	90	no	no	0

[a] presence of the signal peptide indicates that the protein is likely to be secreted

Supplementary Table 15. Genes for ecdysone biosynthesis enzymes identified in the *Schistocerca* embryonic transcriptome.

Gene	Transcript ID	*D. melanogaster* Uniprot ID	e-value	Transcript length (bp)	Protein length (aa)	N-terminus	C-terminus
shade (shd), Cyp314A1	SgreTa0006977	M9PI59	7.19E-143	2603	521	no	yes
shadow (sad), Cyp315A1	SgreTa0006386	Q9VGH1	2.41E-72	1734	481	yes	yes
disembodied (dib), Cyp302A1	SgreTa0014975	Q9NGX9	8.93E-148	1826	526	yes	yes
phantom (phm), Cyp306A1	SgreTd0014875	X2JG03	4.54E-124	1597	482	yes	yes
shroud (sro)	SgreTb0007943	I3VPX6	3.00E-61	1464	357	yes	yes
spook (spo), Cyp307A1	SgreTa0006308	H8F4V5	6.82E-98	2338	303	yes	yes
spook-like	SgreTa0009228	A8Y592	9.09E-48	412	137	no	no
neverland (nvd)	SgreTd0008886	Q1JUZ1	4.46E-93	1567	288	yes	yes
Cyp6t3	nf[a]						
Cyp6u1	SgreTa0011509	A0A0B4LET2	6.74E-05	311	96	yes	no
Cyp303a1	SgreTa0005101	X2JA13	7.54E-141	2633	497	yes	yes
torso	nf						

[a] not found

Supplementary Table 16. RNA-seq differential gene expression of Schistocerca ecdysone biosynthesis enzymes in the pleuropodia at diverse stages.

Day	4			5			6			7		
Gene	**UP/DOWN[a]**	**Fold change**	**Expression[b]**	**UP/DOWN**	**Fold change**	**Expression**	**UP/DOWN**	**Fold change**	**Expression**	**UP/DOWN**	**Fold change**	**Expression**
shade (shd), Cyp314A1	ns[c]			ns			ns			ns		
shadow (sad), Cyp315A1	ns			ns			ns			ns		
disembodied (dib), Cyp302A1	ns			ns			ns			**UP**	**3.03**	**44.57**
phantom (phm), Cyp306A1	ns			ns			ns			ns		
shroud (sro)	ns			ns			ns			DOWN	3.67	
spook (spo), Cyp307A1	ns			**UP**	**49.97**	**17.44**	ns			ns		
spook-like	ns			**UP**	**65.43**	**46.67**	ns			ns		
neverland (nvd)	ns			**UP**	**61.67**	**78.97**	ns			ns		
Cyp6t3 Cyp303a1	ns			ns			**UP**	**503.82**	**81.66**	ns		

8			8-9			10			11		
UP/DOWN	Fold change	Expression	UP/DOWN	Fold change	Expression	UP/DOWN	Fold change	Expression	UP/DOWN	Fold change	Expression
ns			ns			ns			ns		
DOWN	23.44		ns			ns			ns		
UP	**2.92**	**49.05**	ns			**UP**	**4.37**	**52.74**	**UP**	**4.45**	**43.23**
ns			ns			ns			ns		
ns			ns			ns			ns		
ns			ns			ns			ns		
ns			ns			ns			ns		
ns			ns			ns			ns		

12			13		
UP/DOWN	**Fold change**	**Expression**	**UP/DOWN**	**Fold change**	**Expression**
ns			ns		
DOWN	333.62		ns		
UP	**15.03**	**50.91**	**UP**	**43.69**	**71.64**
ns			ns		
ns			ns		
DOWN	17.96		**UP**	**15.96**	**74.05**
DOWN	25.32		ns		
ns			ns		
DOWN	1179.65		DOWN	1647.30	

[a] upregulated (UP)/ downregulated (DOWN)

[b] the DEGs were ranked according to their RPKM (in descending order), the number describes the position of the DEG in the ranked table; top 25% highlighted in black, others in descending level of grey

[c] not significant

Supplementary Table 17. *Schistocerca* genes with GO terms "hormone biosynthetic process" upregulated in the highly secreting pleuropodia.

***S. gregaria* transcript**				**D. melanogaster top hit**	
Transcript ID	**Protein**	**Note**	**Functions in ecdysone biosynthesis**	***D. melanogaster* Uniprot ID**	**e-value**
SgreTa0013987	Juvenile hormone acid O-methyltransferase	methyl transferase, in the corpora allata functions in juvenile hormone biosynthesis		Q9VJK8	1.38E-32
SgreTa0014975	Cytochrome P450 302A1 (dib)	ecdysone biosynthesis in prothoracic glands and other ecdysone producing tissues	x	Q9NGX9	8.93E-148
SgreTa0016782	Carbohydrate sulfotransferase	carbohydrate biosynthetic process		Q9W070	2.20E-40
SgreTa0017764	Uncharacterized Short-chain dehydrogenase-reductase			Q9VDC0	1.23E-46
SgreTb0017908	Niemann Pick type C2 protein homolog (Npc2)	regulates sterol homeostasis and by this also ecdysteroid biosynthesis	x	Q9VQ62	7.98E-35
SgreTa0002115	Dopamine N-acetyltransferase	melatonin biosynthesis		Q94521	1.62E-21
SgreTa0002227	Cytochrome P450 305A1	may be involved in the metabolism of insect hormones by sequence similarity (Uniprot)		Q9VW43	4.39E-107
SgreTa0007915	Juvenile hormone acid O-methyltransferase	methyl transferase, in the corpora allata functions in juvenile hormone biosynthesis		Q9VJK8	1.14E-31

Supplementary Table 18. Sequnces of Primers

Transcript ID	Forward primer (5'-3')	Reverse primer (5'-3')
SgreTa0002695	ATGCCTGGGTGTTGGATAAG	GGAGCATCTATGATGGTCACG
SgreTa0007432	AAGGTTCTTGCAGGATGGTG	AGCTCCACAAATCTGCCTTC
SgreTa0001469	TCATCACTGGCATCTTCTCG	TTTTCACCTCCACGGAGAAC
SgreTa0015941	AACACCGCTACAGGAAATGG	TGCACCTTGAGGTTTGACAG
SgreTa0007802	ATGAGGGCTCTTTGACAACC	ACAGCGCAGACTACGAAATG
SgreTa0005616	GAAGGATTCGCTTACGAAGG	TATCGGGCTCTGGTACTTGC
SgreTa0011044	TGTGAAGGGCCTAGGAAAAG	TCAGTTGCCTTCATCCAGTG
SgreTa0006252	TCCAACACAAAGAGGTGGTG	TGCTGCAGTAAGCAACCAAC
SgreTa0017664	GGACAGAAGACGACACACAGG	ACACGCAGGACAATGAGGAC
SgreTa0005054	TCGGCACACAGAAGTTCAAG	TCCATCGAAGTCGTGCTTTC
SgreTa0002027	ACCCGACATCCTCAAACTTC	TTTGGCTGACTCCCAGAAAC
SgreTa0009118	AGGTATCGCCAAGCACAAAG	GAGTTCTTATCTTGGGGTGCAG
SgreTa0000088	TGTGTCCATTGGATGTCACC	CACATGCTGCTGGATCATTC
SgreTd0002755	GGTCCGGTATTTGGGAAAAC	AACTGAGGTCTCGCACCTTG
SgreTa0014975	TGGATTCCATGTACCAGCAG	TGTCCTTTCAGCCACCTTTC
SgreTa0001341	GGATTCGATCTCAACGCAAG	AGGACAGCGTGTTGTTGTTG
SgreTf0013577	ACGATGCACCAGAACTACCC	TTATTCCCTTCCCGTACAGC
SgreTa0001826	ATGCGTCCATACTTGTGGTG	ATGAACAGCAGCTGGAAAGC
SgreTa0000488	ACCTGTTCTGATGGCGAATC	GCCCCGTCTTCTTTTCTTG
SgreTa0009559	CCCTGAGATTTGGCTTGAAC	CTTCATTTCCTCGTGCCATC
SgreTa0003305	AATGGCTCCAAGACAAGTGG	TCACTTGGAGATGCTGAAGG
SgreTd0003949	TGAGAAGGCAGACGAACATC	AGGGTCAGCAGTGCATTTTC
SgreTb0019973	TCCAGTGATGACACACACACAG	CGAAATGAGGCGAGAGAAAC
SgreTb0006243	CCATGACTTCGCTTTGATCC	TAAGGCTGGTTGAGCACTTC
SgreTa0017707	TTACGTGCGATGTTCGTCAG	AATGGCTGCATAGTCGAAGC
SgreTa0017736	ACTCCTCAACGATGCTTTCG	GTTGCAATCCTTGCGATACC
SgreTa0007897	TCAGGAACTGGGTATGCTTG	TGATCTGGAACAAGCCGTAG
SgreTa0008528	AATTGCCAGGAGTGGATAGG	ATTGTAGGCCAGAGCCAAAC
SgreTa0001449	GGAAAGATTGCTCTGGATGG	ATTCCAAGCTGACCACGAAG
SgreTa0005600	AACTTCCTGCCAGTGGAGAC	AGTGCAGCACATTCAGCTTG
SgreTc0000004	AAGGCCCAGTGTCTGTTTTC	TTTCTCGGGGATGTACTTGG
SgreTc0000003	AGTGCTTTGCCTTGTTGGAC	GTTCACGGAAACGATTGCAC
SgreTa0013453	AAGGCTGCATTGTGGATACC	TGGACGTGAACGATTGTAGC
SgreTa0008219	CAAGTCGAGCAATTCTACGC	TCTCGGGGTTCCATAAGAAG
SgreTd0008886	GGAGCGGTGTTCAAAAAGAC	GAAACAGCCGTGTTCCTTTC
SgreTa0008497	GGAAACAGTGAGGCGAAAAC	AGTTGTTCTGGGCATTAGCC
SgreTd0014875	AGCCCGGACAACACTTCTAC	CCATCATGAGCAGGAACCAG
SgreTa0006386	GACCTCAGCAGCGATCATTC	CACACGCAGGTACATATGAAGG
SgreTa0006977	CTTGCAGATGCAGTCAATGG	TGGCAGTATCTTCCAGAATGG
SgreTa0006308	GTGCATCAAATGCTCACTCG	TGGACGCTAGCACTCTCTAATG
SgreTa0002186	ACTTTTGTGGACCCCTCATC	AGTGGACCAGCCTTTCATAGAC
SgreTf0014307	CAAGATGCCGACTGTGAGTG	GGCGGTAACAGAAACAAAGC

SgreTa0001661	AGGATTGGTCCAGTTTCGTG	TCCATCTCGTCACATCTTCG
SgreTb0016047	ACGTAATTGACAGCCACTCG	ATCGAGTCTTTGGTGGCATC
SgreTa0014626	ATTTACGGCTTGGTCGTAGC	GATGCCGATAGCAAATCCTG
SgreTa0008504	GAGAAATCATCCGGTTGGAG	AAGATGCTGCCCATGATACC
SgreTa0007477	GAGCAGCATTTCCACAAGC	TCATGCGCTTCTCCTTCTG

4. Chapter II - Variation in a pleiotropic regulatory module drives evolution of head shape and eye size in *Drosophila*

The manuscript **'Variation in a pleiotropic regulatory module drives evolution of head shape and eye size in *Drosophila.'*** is the main project of my PhD thesis.

My contributions for this manuscript includes the following parts:

- Conceptualization of project and experiments (together with Dr. Nico Ponsien)
- Bioinformatics analyses (RNA-seq and ATAC-seq)
- Planning and performing experiments (Experimental lab work was supported by Bilen A., Matas de las Heras C., Ayaz S., Niksic A.)
- Data interpretation (together with Dr. Nico Posnien)
- Writing of the first manuscript draft and editing (together with Dr. Nico Posnien)
- Visualization (together with Dr. Nico Posnien)

Contribution of other authors includes:

- Dr. Torres-Oliva, M. and Dr. Almudi, I. generated the transcriptomic dataset
- Prof. Casares, F. provided the analysis and figures for the lineage tracing experiment and the staining of *pnr*-expressing cells during pupal stages.

Status of the manuscript:

In preparation for submission

Title

Variation in a pleiotropic regulatory module drives evolution of head shape and eye size in *Drosophila.*

Authors

Buchberger E.[1], Bilen A.[1], Matas de las Heras C.[1], Ayaz S.[1], Niksic A.[1], Almudi I.[3], Torres-Oliva M.[1,2], Casares F.[3], Posnien N.[1,*]

[1] Universität Göttingen, Department of Developmental Biology, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

[2] Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, University Hospital Schleswig-Holstein, Kiel, Germany

[3] CABD (CSIC/UPO/JA), DMC2 Unit, Pablo de Olavide University campus, Ctra. Utrera km1, 41013 Seville, Spain

* author for correspondence: nposnie@gwdg.de

4.1. Abstract

Insect compound eyes are highly complex organs, which are composed of individual subunits, so called ommatidia. We have recently shown that closely related *Drosophila* species show remarkable differences in eye size and head shape. The eye size differences between *D. melanogaster* and *D. mauritiana* are the result of differences in the number of ommatidia. We use this model to identify the molecular changes underlying the observed morphological variation in adult structures and try to understand how gene regulatory networks (GRNs) in closely related species evolve.

A comparative developmental transcriptomic dataset combined with a transcription factor binding site analysis showed that the GATA factor Pannier (Pnr) regulates many genes that are differentially expressed between *D. melanogaster* and *D. mauritiana* and that the transcript of *pnr* itself is differentially expressed in the two species during eye development. Additionally, we could show that *u-shaped* (*ush),* coding for a co-factor of Pnr, is transcribed and translated in the developing eye-antennal disc. We used the binary GAL4-UAS system and subsequent antibody staining to reveal that the two factors regulate each other. To test, if the regulatory module composed of Pnr and Ush may represent a flexible node in the eye and head developmental GRN, we overexpressed *pnr* and *ush*, respectively in the eye-antennal disc in *D. melanogaster*. We indeed were able to phenocopy aspects of the differences observed between *D. melanogaster* and *D. mauritiana*, showing that higher levels of Pnr lead to a bigger eye area, due to a higher number of ommatidia and a narrower, interstitial face cuticle. In summary, our data suggests that differences in the expression of *pnr* and *ush* might explain part of the variation observed between the head shapes of *D. melanogaster* and *D. mauritiana*.

4.2. Introduction

The capacity of organisms to generate new forms is a key prerequisite for the adaptation to an ever-changing environment. One of the major goals in biological research is to understand the intrinsic and extrinsic forces shaping this morphological variability. Since the genome of an organism contains instructive information about its morphology, generally a first crucial step is the establishment of the genotype-phenotype map for a given morphological trait. The genetic architecture of relatively simple traits has been successfully determined at a high resolution. For instance, natural variation in body pigmentation in the vinegar fly *Drosophila melanogaster* or the beach mouse *Peromyscus polionotus* has been mapped to individual nucleotides affecting the expression of the underlying gene (Jeong et al., 2006) or protein function (Hoekstra et al., 2006), respectively. Also, the genetic basis of the gain or loss of structures like trichomes in *Drosophila* (Arif et al., 2013b; McGregor et al., 2007), pelvic spines in stickleback fish populations (Chan et al., 2010; Xie et al., 2019) or the repeated loss of eyes in cave fish (reviewed in (Krishnan and Rohner, 2016)) has been successfully revealed. However, the genetic changes underlying the evolution of complex traits, such as the size and shape of organs remain largely elusive to date. This is in part due to the polygenic nature of complex quantitative traits. This means that the final observable variation is influenced by many genetic changes with small effect sizes, which are spread throughout various genomic locations, significantly hampering their detection (Boyle et al., 2017; Mackay, 2001). Additionally, quantitative traits are highly context dependent, i.e. time and tissue specific and often influenced by environmental factors like temperature or food availability (e.g. (Casasa and Moczek, 2018; Siomava et al., 2016)). Despite these difficulties, the genetic basis of variation in complex traits has started to be elucidated in recent years. For instance, mandible and craniofacial shape differences between mouse strains are influenced by loci located on most of the chromosomes (Boell et al., 2011; Boell and Tautz, 2011; Burgio et al., 2009; Maga et al., 2015; Pallares et al., 2014). Several studies in *Drosophila* revealed, that loci on several chromosomes underly differences in eye size and head shape (Arif et al., 2013a; Gaspar et al., 2019; Norry and Gomez, 2017). These examples confirm that the genetic architecture of such traits is rather complex and individual causative molecular changes are difficult to determine.

While the genetic architecture of morphological trait variation is being revealed, a mechanistic understanding of the impact of the genomic changes is still missing to date. For

instance, the size and shape of an organ is often exposed to selection pressures at the adult stage when it is functional. However, its appearance is defined during embryonic and post-embryonic development. Therefore, it is conceivable that genetic variation underlying complex trait diversity has a direct impact on its development. The development of an organism and its organs relies on the correct activation and repression of developmental genes which is orchestrated by a complex interplay between gene products in developmental gene regulatory networks (GRNs). These GRNs must be tightly controlled because changes at any node of this network will eventually influence the interaction with its downstream target genes. A balance between a constraint network architecture and flexibility is thus important for allowing changes in size and shape of a certain organ to occur throughout evolution, but at the same time keeping the resulting adult organ fully functional. For the gain or loss of simple morphological traits, a few studies so far have established a clear link between causative genetic variation and GRN architecture. For instance, genetic variation that changes the expression of the zinc finger transcription factor Shavenbaby (Svb) is associated with the presence of trichomes in *Drosophila* larvae (McGregor et al., 2007), while natural variation in adult trichome patterns is explained by genetic variants affecting the expression of the micro RNA *miR-92a* (Arif et al., 2013b). A thorough analysis of the GRNs governing larval and adult trichome development, revealed fundamental differences in the interplay of key developmental regulators (Kittelmann et al., 2018). This data strongly suggests that the GRN architecture poses constraints on the nodes within the network that change during evolution. Due to the polygenic nature of complex morphological traits, the link between genetic variation and GRN architecture is more complicated to establish.

A typical approach to address this gap could be to first identify genetic variants associated with morphological diversity and place the candidates into the GRN context in a second step. As an alternative, we propose here to first identify putative flexible nodes within the GRN governing the development of a variable morphological trait. We suggest that the data obtained from this first step can subsequently be used in follow-up studies to reveal the causative genetic variation associated with trait variation. To identify 'flexible nodes' in an otherwise constraint GRN, we studied genome wide patterns of developmental gene expression variation. We assume that flexible nodes can be identified by their effects on downstream target genes. As model we compared eye and head development in the two closely related *Drosophila* species *D. melanogaster* and *D. mauritiana*, which vary extensively

in adult eye size and head shape (Posnien et al., 2012). It has recently been shown that *D. mauritiana* develops larger compound eyes due to a higher number of individual ommatidia especially in the dorsal eye. The bigger eyes of *D. mauritiana* are accompanied by a narrower interstitial head cuticle (Posnien et al., 2012), recapitulating the common origin of eye and head cuticle tissue from the same eye-antennal imaginal disc during larval development (Haynie and Bryant, 1986). Since the GRN governing eye-antennal disc development is extensively studied and well understood in *D. melanogaster* (Kumar, 2009; Potier et al., 2014; Treisman, 2013), this process represents an excellent model to link morphological diversification to developmental and genetic variation.

We applied RNA-seq at different developmental stages of eye-antennal discs in *D. melanogaster* and *D. mauritiana*. A systematic co-expression and transcription factor enrichment analysis revealed that many differentially expressed genes were regulated by the GATA transcription factor Pannier (Pnr). Our results suggest that Pnr plays a dual role in the underlying GRN since it activates and represses its target genes. The repressive role is most likely mediated by its co-factor U-shaped (Ush) which is, in contrast to previous reports, co-expressed with Pnr during eye-antennal disc development. We applied functional genetics approaches to establish that Ush and Pnr interact genetically during eye-antennal disc development and are thus involved in the same regulatory module. Finally, we show quantitative expression differences of *pnr* and *ush* between *D. melanogaster* and *D. mauritiana* and that the overexpression of *pnr* in *D. melanogaster* phenocopies aspects of the *D. mauritiana* like head shape and eye size. Our data confirms a role of Pnr in morphological differences observed between *D. melanogaster* and *D. mauritiana* and therefore suggest that Pnr might be one flexible node in the conserved eye-antennal GRN.

4.3. Results

4.3.1. *Drosophila melanogaster* and *D. mauritiana* exhibit differences in dorsal head shape

Eye size and head shape vary extensively between *Drosophila melanogaster* and *D. mauritiana* with the latter having bigger eyes due to more ommatidia at the expense of interstitial face cuticle (Arif et al., 2013a; Hilbrant et al., 2014; Posnien et al., 2012). Since eye size differences are most pronounced in the dorsal part (Posnien et al., 2012), we proposed that the shape of the dorsal interstitial cuticle may vary as well. To test this hypothesis, we

comprehensively quantified differences in the dorsal head morphology among the two sister species.

We placed 57 landmarks on pictures of dorsal heads (Figure 17A) covering the main dorsal cuticle regions (Figure 17A, (Haynie and Bryant, 1986)) and we applied a geometric morphometrics analysis. A discriminate function analysis clearly distinguished the head shapes of *D. melanogaster* and *D. mauritiana* (Figure 17B). In accordance with previous data (Posnien et al., 2012), we found main differences in dorsal eye size with the eye area protruding more towards the back of the head in *D. mauritiana* (Figure 17B). The posterior expansion of the eye area in *D. mauritiana* was accompanied by a narrower dorsal head region, which affected both the orbital cuticle (OC) and the dorsal frons (DF) region (compare to Figure 17A). The ocellar complex was slightly shifted ventrally. In *D. melanogaster*, the eye area was clearly smaller, whereas both dorsal head regions (OC and DF) were larger and the ocellar complex was shifted dorsally (Figure 17B).

In summary, we found that *D. melanogaster* and *D. mauritiana* do not only differ in the size of the dorsal eye area, but also exhibit variation in the relative contribution of different head regions to the dorsal head capsule.

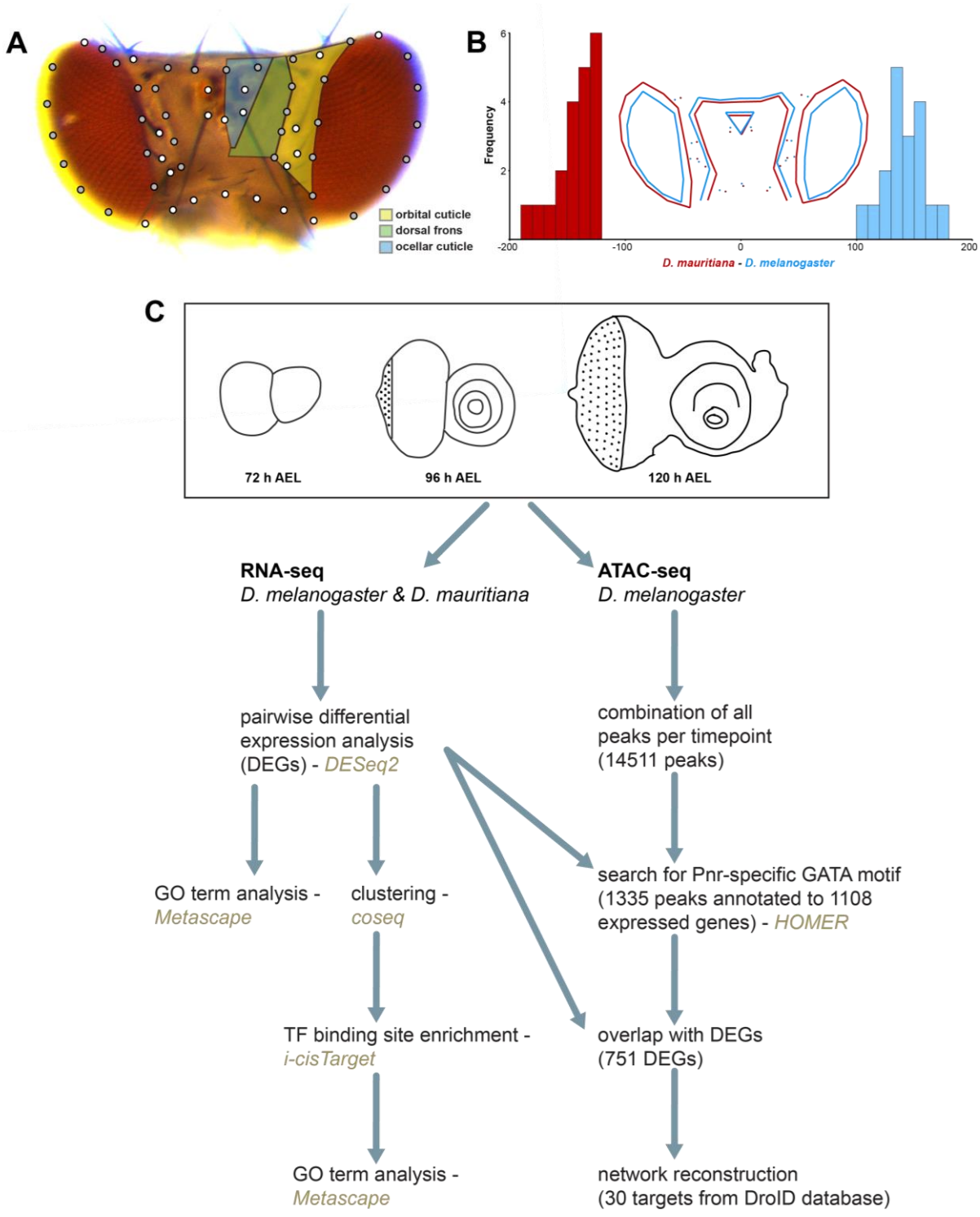


Figure 17. A. Dorsal view of a *Drosophila* head and schematic representation of the dorsal head structures in *Drosophila*. The dorsal *Drosophila* head cuticle consists of three morphologically distinguishable regions, namely the orbital cuticle next to the compound eye (yellow), the dorsal frons (green) and the ocellar cuticle (blue). The dots show the 57 landmarks that were used to analyze head shape, where the white landmarks represent fixed landmarks and the grey ones represent sliding landmarks. **B.** Mean head shape of *D. melanogaster* (blue) and *D. mauritiana* (red) after discriminate function analysis, which clearly distinguished the two groups based on their dorsal head shapes. **C.** Experimental setup of the bioinformatics analysis. Arrows point to each step in the pipeline; Left side: Transcriptomic datasets were generated for developing eye-antennal discs in both species at three developmental stages, namely 72h AEL (after egg laying; late L2), 96h AEL (mid L3) and 120h AEL (late L3). The scheme shows the workflow from data generation to, differential expression analysis to clustering of the

differentially expressed genes. Right side: An ATAC-seq dataset was generated for developing eye-antennal discs in *D. melanogaster* at the same three stages. We defined a list of potential Pnr target genes, using motif search in open chromatin sequences and combined this approach with data from the DroID database, to reconstruct the close network around the GATA-factor.

4.3.2. Difference in the transcriptomics landscape recapitulate observed morphological differences between *D. melanogaster* and *D. mauritiana*

To reveal the molecular basis of the size and shape differences in dorsal head structures, we obtained comparative transcriptomes for three stages of eye-antennal discs development. The stages represented the onset of differentiation (72 h AEL), the progression (96 h AEL) and termination of differentiation (120 h AEL), respectively (Figure 17C) (Torres-Oliva et al., 2018).

A global analysis of the expression data showed that 72 % of variation in the dataset was due to differences between 72h and 96h AEL (Supplementary Figure 8). This observation was confirmed by a pairwise differential expression analysis to determine the number of genes that were differentially expressed between *D. melanogaster* and *D. mauritiana* for each developmental stage. At 72h AEL we found the highest number of differentially expressed genes (DEGs), namely 6,683. This number decreased in later stages with 3,260 and 2,380 DEGs at 96h AEL and 120 h AEL, respectively (Supplementary Figure 9A). We did not find a biased expression difference between species since we observed a more or less equal number of DEGs with higher expression in *D. melanogaster* and *D. mauritiana*, respectively (Supplementary Figure 9A). To test whether the DEGs may be enriched for genes with specific cellular or molecular functions, we performed a gene ontology (GO) enrichment analysis. Indeed, we saw that stage specific DEGs are enriched in GO categories that can recapitulate the cellular events that happen at each respective stage. At 72h AEL we found DEGs upregulated in *D. mauritiana* and enriched in establishment and maintenance of cell polarity, a process which is highly important for overall disc growth and the mirror arrangement of the future ommatidia (e.g. (Jenny, 2010)). Also, DEGs were enriched in signal transduction pathways, for instance protein kinase A signalling (e.g. (Chanut and Heberlein, n.d.; Domínguez, 1999; Pan and Rubin, 1995; Strutt et al., 1995)), Inositol phosphate metabolism (e.g. (Seeds et al., 2015; Tsui and York, 2010)), and TORC signalling (e.g. (Wang and Huang, 2009)), all of which have shown to be involved in *Drosophila* eye development. In *D. melanogaster*, genes were predominantly enriched in cell cycle processes, consistent with the proliferation going on during this early stage (Casares and Almudi, 2016; Kenyon et al., 2003) (Supplementary Figure 9B). At 96h AEL differentiation events with more specific functions related to neural and photoreceptor

development were captured using GO enrichment analysis, like R7 cell development and neural-related GO terms, reflecting the onset of the morphogenetic furrow at this time point, leaving behind differentiated ommatidia (Bonini and Choi, 1995; Heberlein and Moses, 1995; Treisman and Heberlein, 1998) (Supplementary Figure 9C). At 120h AEL we found, among many genes involved in metabolic pathways, differences and genes important for cuticle development. Overall, we were able to recapitulate the differences we observe in the adult flies already in the developing larval tissue (Supplementary Figure 9D).

Overall, we found a substantial number of DEGs between *D. mauritiana* and *D. melanogaster* during eye-antennal disc development, suggesting that we were able to recapitulate the differences we observe in the adult flies already in the developing larval tissue. The observation that these DEGs are involved in crucial developmental processes and molecular pathways suggests that various developmental mechanisms may contribute to morphological diversification between species. Also, this vast range of processes clearly reflects the development of various head regions and sensory organs from one single tissue.

4.3.3. Central transcription factors regulate differentially expressed genes

Since genes involved in central developmental processes are differentially expressed between *D. melanogaster* and *D. mauritiana,* we hypothesized that also key transcriptional regulators may be involved in their differential regulation. To get a global overview of differential gene expression dynamics across both species and time points, we clustered all genes that were differentially expressed in at least one stage according to their expression dynamics. This analysis resulted in 15 unique clusters based on 8,350 genes. Each cluster thus contained genes that share expression profiles across species and developmental stages. A gene ontology (GO) enrichment analysis supported the specificity of the clustering approach (Supplementary Figure 10).

Assuming that co-expressed genes could be regulated by the same transcription factors, we identified putative shared transcription factor binding sites enriched in the regulatory regions of genes present in each expression cluster (see Materials and Methods for details). The unique expression dynamics of each cluster was recapitulated by a specific set of transcription factors involved in the regulation of genes in each cluster (Supplementary Figure 10). Among the enriched motifs, we found binding sites for transcription factors which have previously been described to be involved in eye-antennal disc development. For instance, in

cluster 6 and 7 we found motifs for Lola that regulates ocelli (Mishra et al., 2016), photoreceptor and cone cell development (Zheng and Carthew, 2008). In cluster 10 we found motifs for Blimp-1, a transcriptional repressor associated with Ecdysone signalling (Neto et al., 2017), that has been shown to control the progression of the morphogenetic furrow and thus differentiation in the eye-antennal disc (Brennan et al., 1998). Intriguingly, genes in the same cluster were enriched for Ecdysone receptor (EcR) motifs, further supporting the cooperation of Blimp-1 and Ecdysone signalling (Agawa et al., 2007; Akagi and Ueda, 2011). Transcription factors that have been shown to be involved in photoreceptor development were enriched in cluster 11. For instance, Nejire (Nej) is involved in determination of photoreceptor cell fate (Kumar et al., 2004) and Jun-related antigen (Jra), a member of the c-Jun N-terminal kinase (JNK) pathway, is involved in establishment cell polarity and R3/R4 photoreceptor development (Ciapponi, 2001; Mlodzik, 2002; Weber et al., 2000). In cluster 15 we found an enrichment for the binding sites of Tramtrack (Ttk), a transcriptional repressor that negatively influences the Epidermal growth factor receptor (EGFR) signalling pathway in the eye-antennal disc (Kumar and Moses, 2001). Additionally, Ttk is involved in cone cell (Shi and Noll, 2009) photoreceptor development (Xiong and Montell, 1993). A strong enrichment of GATA motifs was observed in clusters 2, 3, 5 and 8. Motifs of the GATA transcription factor Pnr, that is playing a role in the establishment of the early dorsal ventral axis of the eye and later dorsal head development (Maurel-Zaffran and Treisman, 2000; Singh and Choi, 2003) were enriched in all four clusters with the strongest enrichment in cluster 8.

Intriguingly, 12 of the 20 identified transcription factors (60%) were also differentially expressed between *D. melanogaster* and *D. mauritiana* (labelled transcription factors in Supplementary Figure 10), suggesting that variation in expression of these central regulators had a major impact on the transcriptomics landscape of developing eye-antennal discs among species.

In summary, we could show that interspecific variation in expression of central transcription factors very likely drive the differential expression of a high number of target genes which control important developmental processes during eye-antennal disc development.

4.3.4. Pannier regulates genes that are differentially expressed between *D. melanogaster* and *D. mauritiana*

Pnr is an interesting candidate transcription factor that may be involved in the development of differences in dorsal head morphology as well as eye size observed between *D. melanogaster* and *D. mauritiana* for the following reasons: 1. Our global clustering and motif enrichment analyses suggest that Pnr regulates many DEGs between both species. 2. *pnr* itself is differentially expressed between *D. melanogaster* and *D. mauritiana*. 3. Pnr is known to be expressed in the dorsal portion of the eye-antennal disc (Maurel-Zaffran and Treisman, 2000; and see below) and it determines the dorsal-ventral axis of the retinal field in the early L2 discs (Maurel-Zaffran and Treisman, 2000; Singh et al., 2005; Singh and Choi, 2003). Additionally, later during eye-antennal disc development, Pnr influences the ratio of retinal and head cuticle fate in the dorsal disc by repressing retinal determination genes (Oros et al., 2010). Therefore, we sought to validate and refine our global differential expression data focusing on Pnr.

First, we asked at what stage of eye-antennal disc development *pnr* was differentially expressed between species. Based on our transcriptomic dataset we found significantly higher expression in *D. mauritiana* at 120h AEL (Figure 18A). This trend was further confirmed by real-time qPCR (Supplementary Figure 11).

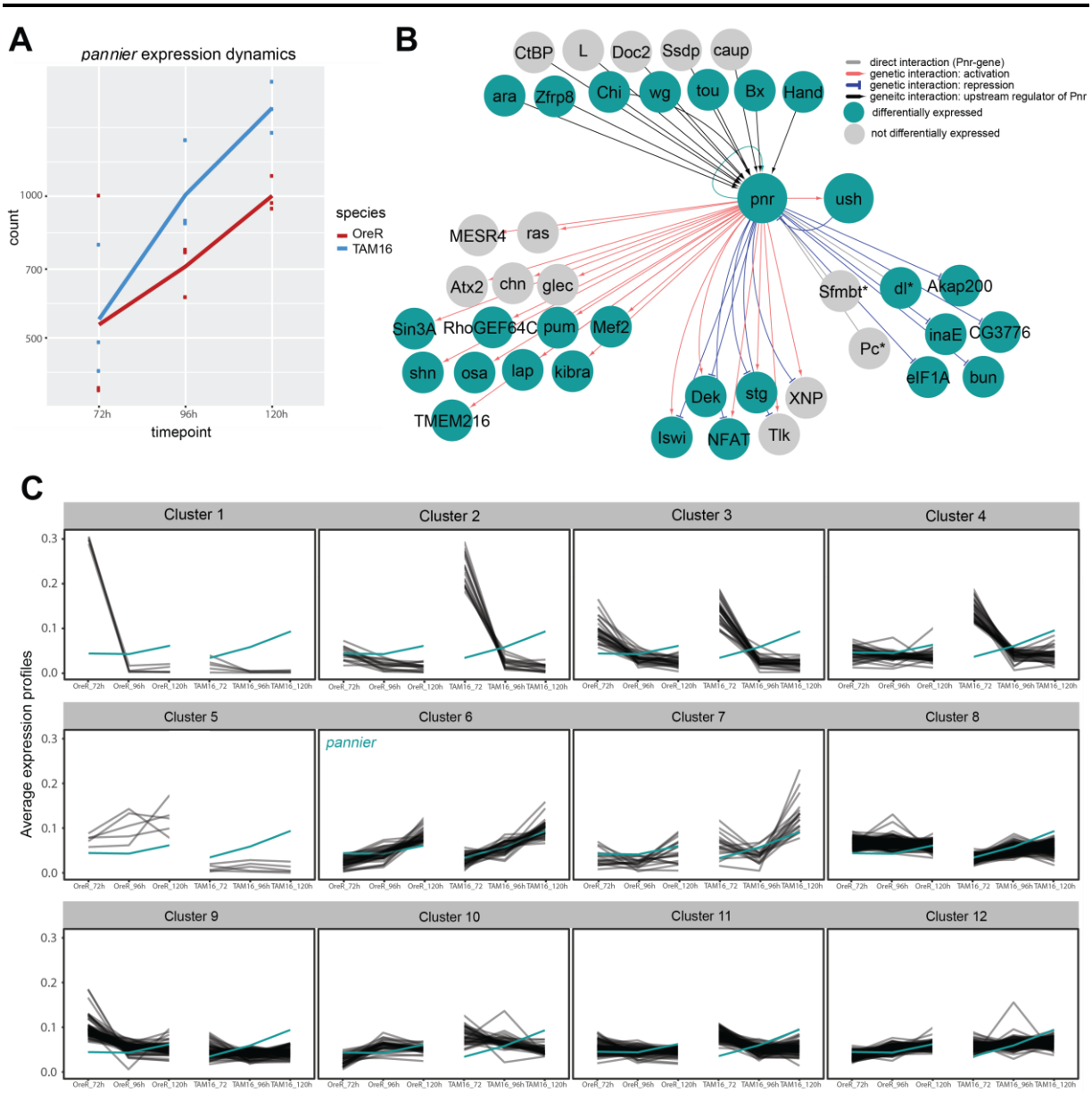


Figure 18. A. Expression dynamics of the *pnr* transcript at the three developmental stages in *D. melanogaster* (red) and *D. mauritiana* (blue). **B.** Network reconstruction of known interactions upstream and downstream targets of Pnr (DroID (Yu et al., 2008)) that overlap with our Pnr target gene list. Cyan nodes represent target genes that are differentially expressed between *D. melanogaster* and *D. mauritiana* in at least one of the three studied developmental stages. Grey nodes represent predicted targets of Pnr based on our target gene list but are not differentially expressed. Black edges describe potential upstream regulators of Pnr based on DroID. Red arrows point towards Pnr target genes that are annotated as being 'activated' by Pnr in DroID, whereas blue edges point to genes where the interaction between Pnr and the gene is annotated as 'repressing'. Grey edges describe interactions that are annotated as direct TF-gene interactions in DroID. **C.** Hierarchical clustering of read counts of predicted Pnr target genes (based on our target gene list) which were found to be differentially expressed in at least one developmental stage. The cyan line in each cluster represents *pnr* expression, which itself is a member of Cluster 6. Left side of each cluster: Expression dynamics of genes in *D. melanogaster* (OreR), Right side of each cluster: Expression dynamics of genes in *D. mauritiana* (TAM16).

Next, we wanted to define a list of putative direct Pnr target genes. This was crucial since the database used to infer motif enrichment was based on ChIP-Chip and ChIP-seq

experiments that were not conducted in *Drosophila* eye-antennal discs (Herrmann et al., 2012; Imrichová et al., 2015). To obtain tissue and stage specific putative target genes, we assessed accessible chromatin regions by generating an ATAC-seq dataset for *D. melanogaster* eye-antennal discs covering the same three time points used for the transcriptomic dataset (Figure 17C). We found 14,511 unique peaks across all three timepoints. In the open chromatin regions, we revealed 1,335 Pnr-specific GATA motifs associated with 1108 genes expressed in our RNA-seq dataset (see Materials and Methods for details), suggesting that they were active during eye and head development. A cross validation of the putative Pnr target genes using the i-*cis*Target tool confirmed an enrichment for Pnr, Nej, pMad and Mef2 binding sites (Supplementary Figure 12A). The identification of putative pMad target genes among Pnr targets may recapitulate the previous observation that both proteins interact physically during larval development (Kim et al., 2017). The putative Pnr target genes were highly enriched in GO terms like signal transduction, development, growth and cell cycle progression as well as in very specific terms such as compound eye development (Supplementary Figure 12B), recapitulating known functions of Pnr during eye-antennal disc development.

We further assessed the reliability of our target gene identification by searching for known target genes of Pnr. Among the putative target genes, we found *Angiotensin-converting enzyme* (*Ance*) (Supplementary Table 19), which is regulated by Pannier and pMad during *Drosophila* larval development (Kim et al., 2017). *pnr* itself is autoregulated in the wing imaginal disc (Fromental-Ramain et al., 2010, 2008). Accordingly, we found *pnr* as target gene as well (Figure 18C, Supplementary Table 19). We did not find *wg* as putative target gene, which is consistent with the study of Pereira and collegues, who suggested that Pnr does not activate wg expression in the peripodial membrane (Pereira et al., 2006). Conserved GATA motifs, that were though shown to be not responsive to Pnr bining (Pereira et al., 2006) lie indeed between significantly called peaks of a highly accessible intergenic region between the *wg* and *wg6* loci (Supplementary Figure 13). Overall, we were able to obtain a high confidence Pnr target gene list.

Our initial cluster analysis suggested that Pnr may regulate many genes that are differentially expressed between *D. melanogaster* and *D. mauritiana.* We could confirm this observation because 67.8 % (751 of the 1,108) of the expressed target genes showed expression differences between *D. melanogaster* and *D. mauritiana* in at least one stage.

In summary, we could show that *pnr* expression was significantly higher in *D. mauritiana* at 120h AEL and we identified a list of high confidence Pnr target genes which are mainly involved in signalling and developmental processes, cell cycle progression and growth. Most of the Pnr target genes were differentially expressed between *D. melanogaster* and *D. mauritiana*.

4.3.5. Pnr activates and represses target genes in the eye-antennal disc

To gain more detailed insights into the expression dynamics throughout eye-antennal disc development, we next clustered the differentially expressed Pnr target genes according to their expression profiles (Figure 18C). Among the 12 obtained clusters, we found *pnr* itself in cluster 6. While the other genes in cluster 6 as well as genes in clusters 7, 8 and 12 showed a similar expression dynamics as *pnr*, we also found clusters in which the expression of the target genes showed the exact opposite trend. For instance, the Pnr target genes in cluster 3 were highly expressed at 72 h AEL in *D. mauritiana*, while *pnr* itself showed a relatively low expression (Figure 18C). The expression of the same target genes decreased at 120h AEL with *pnr* expression increasing at the same time. This contrasting expression profile suggests that those target genes may be repressed by Pnr action. In contrast, genes in clusters that show the same dynamics as *pnr* may be positively regulated by Pnr.

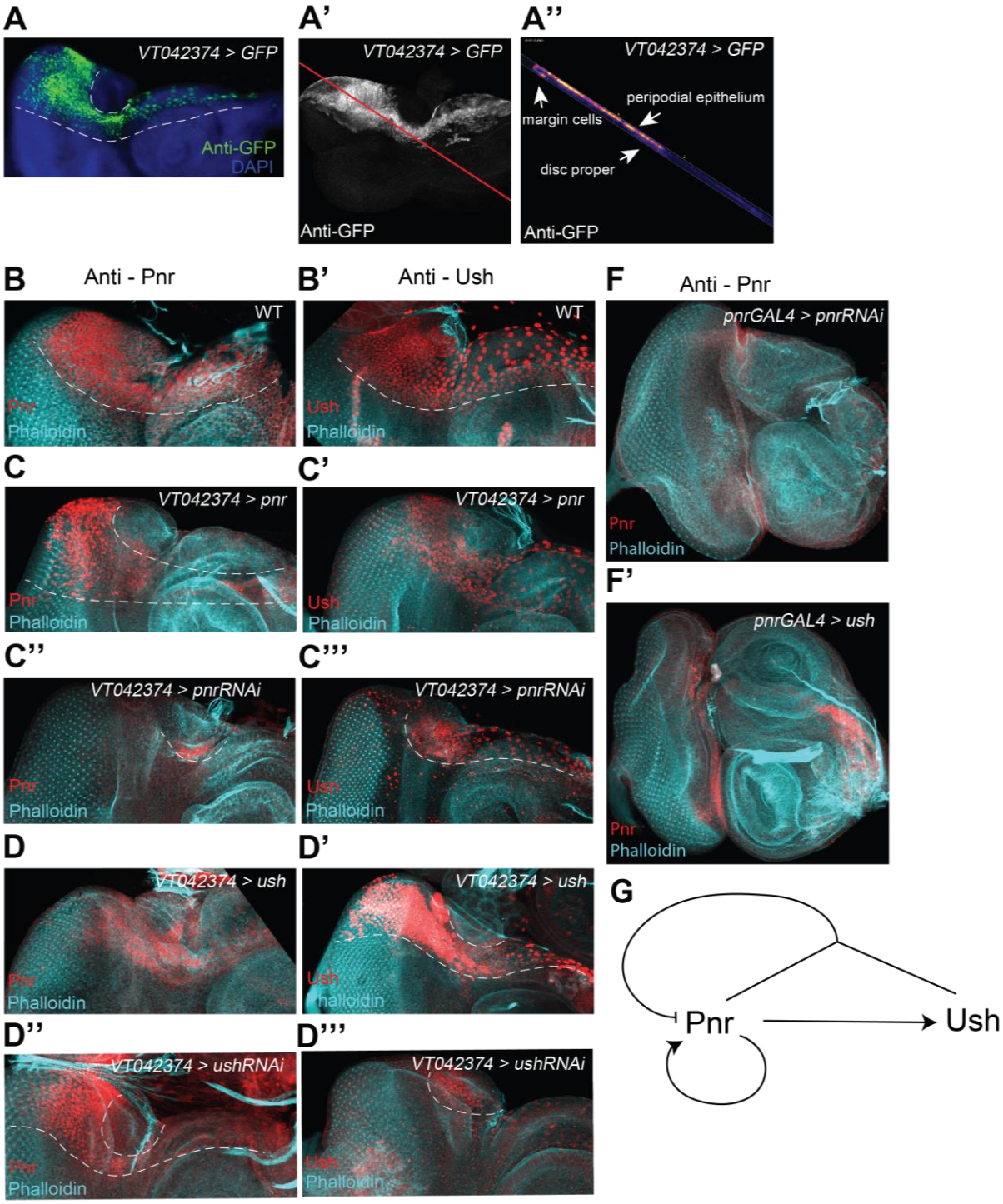


Figure 19 A. *VT042374* drives expression in the dorsal part of the developing eye-antennal disc being partially reminiscent of the endogenous *pnr* expression. **A'.** Vertical section of *VT042374*>GFP. *VT042374* drives predominantly in the cells of the peripodial epithelium and in a few cells of the margin cells, which connect the peripodial epithelium with the disc proper. **B.** Pnr is localized in the dorsal part of the developing eye-antennal disc of *D. melanogaster* (detected with α-Pnr antibody). White, dotted lines mark the area where antibody staining could be detected. Phalloidin (in cyan) was used to show the structures of the eye-antennal discs. **B'.** Ush was detected in the same dorsal region of the eye-antennal disc (detected with α-Ush antibody). **C – C''': Overexpression and knock-down of *pnr*. C.** Pnr localization after overexpression of *pnr* using the *VT042374* driver line. **C'.** Ush localization after overexpression of *pnr* using the *VT042374* driver line. **C''.** Pnr localization after knock-down of *pnr* using the *VT042374* driver line and the *pnrRNAi2* effector line. **C'''.** Ush localization after knock-down of *pnr* using the *VT042374* driver line and the *pnrRNAi2* effector line. **D-D''': Overexpression and**

knock-down of *ush*. D. Pnr localization after overexpression of *ush* using the *VT042374* driver line. **D'.** Ush localization after overexpression of *ush* using the *VT042374* driver line. **D''.** Pnr localization after knock-down of *ush* using the *VT042374* driver line. **D'''.** Ush localization after knock-down of *ush* using the *VT042374* driver line. **F,F'.** Overexpression of *ush* using *pnr*GAL4 (**F'**) recapitulates knock-down of *pnr* (**F**), resulting in duplication of the antennal part of the disc. Pnr is only detectable in a few remaining cells. **G.** Proposed model of how Pnr and its co-factor Ush interact in the developing eye-antennal disc.

To get a clearer picture of whether Pnr may indeed be involved in activation and repression of target genes, we integrated known interactions from the DroID interaction database (Yu et al., 2008). We selected all known target genes of Pnr from this database and overlapped them with our list of putative Pnr target genes. We found three target genes in our list for which the direct interaction of Pnr and the target genes (i.e. Pnr-regulatory sequence interaction) was already shown (*dl, Pc* and *Sfmbt*). Additionally, we found 25 of our high confidence target genes in the list of known genetic interactions (Figure 18B). The fact that we found GATA motifs in the putative regulatory regions of these genes, suggests that they might be direct Pnr target genes. Since the DroID database contains the information, whether interactions are "suppressible" or "enhanceable", we tested if target genes of both categories were present in our dataset. Indeed, 14 of the 29 target genes showed "enhanceable" and 8 showed "suppressible" interactions with Pnr, respectively. 6 target genes showed both types of interactions. Intriguingly, 21 of the 30 putative Pnr target genes (68%) found in the DroID database were differentially expressed (Figure 18B).

The clustering analysis of differentially expressed target genes suggests that Pnr activates and represses its targets in the eye-antennal disc. An in-depth analysis of previously known interactions strongly supported that Pnr target genes are under positive as well as negative transcriptional control.

4.3.6. Pannier and its co-repressor U-shaped participate in the same regulatory network during eye- and head development in *Drosophila*

Our observation and previous reports of a dual regulatory role of Pnr during eye-antennal disc development may be mediated by the presence of a co-factor that modulates its regulatory role. In the developing wing imaginal disc, it has been shown that Pnr acquires a repressing mode of regulation upon heterodimerization with its co-factor U-shaped (Ush) (Fossett et al., 2001; Haenlin et al., 1997; Sorrentino et al., 2007). It has previously been stated that Ush is not expressed in the eye-antennal disc (Fossett et al., 2001; Maurel-Zaffran and Treisman, 2000). However, in our RNA-seq data we found the transcript of *ush* being expressed

during eye-antennal disc development (Supplementary Figure 14A). Therefore, we hypothesized that Ush may act as a co-factor in this tissue.

A role of Ush as co-factor of Pnr requires both proteins to be present in the same cells of the eye-antennal disc. Since *pnr* expression in the eye-antennal disc has only been studied based on Gal4 driver lines, we first characterized the localization of Pnr using a newly generated antibody. We found that the protein is located, as previously reported, in the large nuclei of the dorsal peripodial epithelium (Figure 19B). Additionally, Pnr was detected in a few cell rows in the disc proper, most probably in a subset of the cuboidal margin cells (Supplementary Figure 15B-B''). In later stages, the Pnr staining was less intense in the future ocellar complex region (Supplementary Figure 15AA''). Lineage tracing experiments showed that descendants of *pnr*-positive cells extend further ventrally in the peripodial epithelium (Supplementary Figure 15C-C'''). Additionally, we observed descendants of *pnr*-positive cells in the dorsal disc margin as well as in the disc proper (Supplementary Figure 15D-D'''). Using a newly generated antibody against Ush, we confirmed the presence of the Ush protein during eye-antennal disc development in *D. melanogaster* (Figure 19B'). As shown for Pnr, the Ush protein is localized in the nuclei of the peripodial epithelium in the dorsal part of the eye-antennal disc, spanning the antennal, the ocellar and parts of the future head cuticle regions (Figure 19B'). We also observed Ush expression in potential adjacent cuboidal margin cells (Supplementary Figure 16). Therefore, Ush and Pnr expression largely overlaps in the dorsal region of the eye-antennal disc (see also Supplementary Figure 14B), suggesting that they could indeed interact in the developing head. Please note that Ush is not only expressed during eye-antennal disc development, but also necessary for proper head development. Knockdown of *ush* in the dorsal developing eye-antennal disc consistently led to the loss of posterior vertical bristles (pVT – (Chyb and Gompel, 2013)), and irregularities at the border of orbital cuticle and dorsal frons (Supplementary Figure 14C), while the upregulation of *ush* affected the overall head shape and loss or gain of the pVT and adjacent bristles (Supplementary Figure 14C'). The effect on bristle patterns is consistent with the reported role of Ush in bristle formation on the thorax (Cubadda et al., 1997; Haenlin et al., 1997).

The co-expression of Pnr and Ush suggested that both genes may interact genetically. To test this, we assessed the effect of gain- and loss of function of both genes on each other using the binary GAL4-UAS system in combination with Immunohistology. Since we aimed at

modulating the expression of both genes within the endogenous domains, we used GAL4 driver lines, which drive expression in different dorsal regions of the developing eye-antennal disc. *VT042374* activity was reminiscent of the *pnr* expression domain in the peripodial epithelium and in marginal cells except for a small region in the presumptive ocelli domain (Figure 19A). Regulatory elements of this line overlap with two open chromatin ATAC-seq peaks in an intronic region of the *pnr* locus (Supplementary Figure 17), suggesting that indeed partial endogenous *pnr* expression is reported. Additionally, we used the *oc*-GAL4 driver line that drove expression in the ocellar complex region that was not covered by the *VT042374* line (Supplementary Figure 18 A-A'').

Knock-down of *pnr* in the eye-antennal disc using *VT042374* led to depletion of both, Pnr protein and Ush protein (Figure 19C''-C'''). This finding showed on the one hand that the *pnr* knock-down worked efficiently and suggests on the other hand that Pnr is necessary for the expression of *ush*. Note that this result could also be observed using the *oc*-GAL4 driver line, where even though the discs show great deformation after *pnr* knockdown, the Ush protein was clearly detected only in a smaller region (Supplementary Figure 18D'). Pnr was upregulated upon overexpression using the *VT042374* driver (Figure 19C) and the *oc* driver (Supplementary Figure 18C). While the effect on Ush was not obvious after *pnr* overexpression using the *VT042374* driver (Figure 19C'), slight upregulation was observed when the *oc* driver was used (Supplementary Figure 18C').

The knockdown of *ush* using the *VT042374* driver line resulted in a complete loss of Ush protein in the expected region (Figure 19D'''), confirming that the knock-down worked efficiently. Conversely, we observed upregulation of Pnr expression in the region where RNAi against *ush* was driven (Figure 19D''), suggesting that the presence of Ush results in *pnr* repression. Overexpression of *ush* using the *VT042374* driver line resulted in a reduction of Pnr expression (Figure 19D). To confirm this observation, we made use of a previously reported double antenna phenotype upon loss of Pnr function (Oros et al., 2010), that we also found after *pnr* RNAi (Figure 19F). Intriguingly, overexpression of *ush* using a stronger *pnr* driver line (*pnr*-GAL4, (Fossett et al., 2001; Heitzler et al., 1996)) resulted in the same double antenna phenotype (Figure 19F'), supporting the observation that Ush is involved in repression of *pnr* expression.

In summary, we could show that Ush and Pnr are spatially co-expressed during eye-antennal disc development. Our gain- and loss of function experiments showed that Ush is necessary for proper head development. Furthermore, we found evidence for genetic interactions between Ush and Pnr during eye-antennal disc development (Figure 19G), implying that both participate in the same regulatory network.

4.3.7. Overexpression of *pannier* phenocopies aspects of the differences observed between *D. melanogaster* and *D. mauritiana*.

The findings obtained so far strongly suggest that Pnr and Ush may contribute to the morphological differences observed between *D. melanogaster* and *D. mauritiana* adult heads: 1) *pnr* and *ush* showed higher expression in *D. mauritiana* during eye-antennal disc development. 2) Both genes are expressed in the dorsal region of the disc and they cross-regulate each other. 3) Many target genes of Pnr are differentially expressed between both species. To test if changes in *pnr* expression indeed have the potential to explain naturally occurring differences in eye size and head shape we quantitatively analysed the shape of fly heads originating from gain- and loss of function experiments.

We crossed the *VT042374* driver line to a UAS-*pnr* overexpression line to mimic higher *pnr* expression in *D. melanogaster* as observed in *D. mauritiana*. Additionally, we crossed the *VT042374* line to two UAS-*pnr*RNAi lines. Overexpression of *pnr* led to a duplication of one of the posterior vertical bristles at the eye rim, while *pnr* RNAi resulted in a loss of bristles (Figure 20A). This observation is consistent with the reported role of Pnr in governing bristle pattern formation (Heitzler et al., 1996) and confirms the specificity of the performed gain- and loss of function experiments.

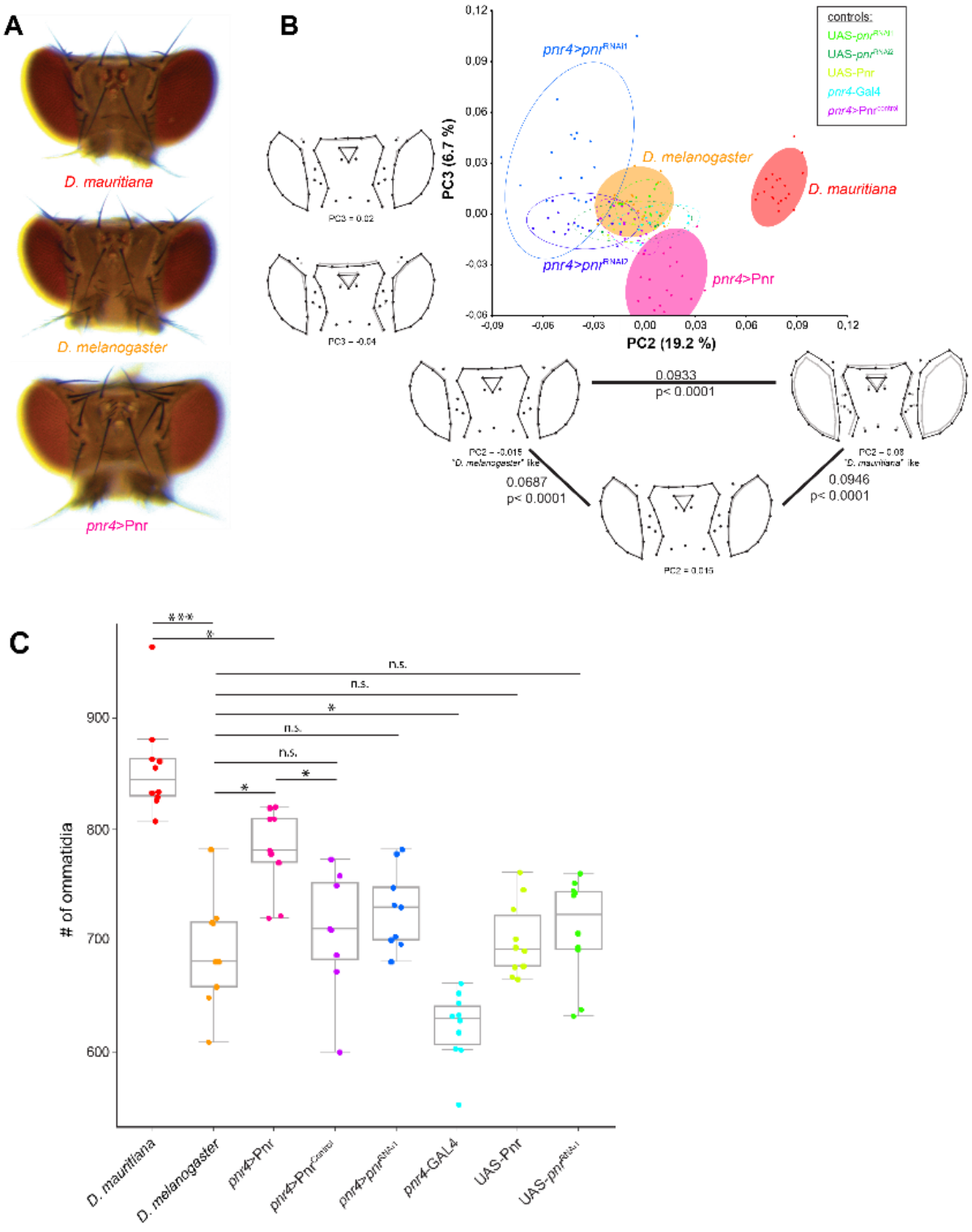


Figure 20. A. Dorsal view of heads of *D. mauritiana* (top), *D. melanogaster* (middle) and *VT042374* ('*pnr4*') >*pnr* (*pnr* overexpression) flies. **B.** Principle component analysis of dorsal head shapes. Shown are PC2 against PC3. Red and orange clouds represent the 'WT-like' head shapes (*D. melanogaster* in orange and *D. mauritiana* in red). Overexpression of *pnr* is represented in pink. The blue empty circles represent knock-down of *pnr*, with a weak effector line in dark blue and a strong effector RNAi line in light blue. The dotted lined circles represent head shapes of the parental UAS- and GAL4 fly lines, respectively, that were used to set up overexpression and knock-down of *pnr*. Extracting Procrustes distances between the groups showed that head shapes of *D. melanogaster*, *D. mauritiana* and flies upon *pnr* overexpression are all significantly different from each other. **C.**

Boxplot of ommatidia numbers in each of the lines (same color-code as in **B**.). Statistical comparisons represent pair-wise comparisons after Tukey HSD test: *** p<0.0001; *p<0.05.

Apart from extra setae at the rim of the eye, overexpression of *pnr* in the dorsal head region did not result in major morphological perturbations (Figure 20A). To quantitatively compare head shapes, we applied geometric morphometrics based on 57 landmarks placed on the dorsal head pictures. A principal component analysis showed that 40.9% of the observed variation in head shape could be assigned to technical artefacts related to the positioning of the heads (PC1, Supplementary Figure 19). Therefore, we excluded the first principal component (PC1) and analysed PC2 and PC3 in more detail. PC2 explained 19.2% of the observed variation in the head shape dataset and PC3 explained 6.7% (Figure 20B). Variation along PC2 mainly captured differences in the proportion of eye vs. cuticle tissue in the dorsal head, as well as the location of the ocellar region. PC3 explained mostly differences in the dorsal-posterior head cuticle and the location of the ocellar region (Figure 20B). The overexpression of *pnr* in the dorsal head region resulted in a shift from a "*D. melanogaster*"-like shape towards a more "*D. mauritiana*"-like shape along PC2. The shape analysis revealed an enlargement of the eyes in the dorsal head region that was accompanied by a slight reduction of the head cuticle between the two eyes (Figure 20B). Ommatidia counting in entire eyes confirmed that the increase in eye area upon *pnr* overexpression observed in our shape analysis was indeed due an increase in number of ommatidia (Figure 20C). Note that *pnr* RNAi influenced overall head shape (Figure 20B), but no impact on the number of ommatidia was observed (Figure 20C).

We also observed that the occipital region of the posterior head was more convex upon overexpression (Supplementary Figure 20A), whereas downregulation consistently led to an enlargement of these regions (Supplementary Figure 20B). To test, whether the occipital region also showed differences between *D. melangaster* and *D. mauritiana*, we performed a shape analysis with additional landmarks. Intriguingly, the occipital region was clearly convex in *D. mauritiana* and more concave in *D. melanogaster* (Supplementary Figure 20C-E). Detection of *pnr* expression in pupae stages (Supplementary Figure 20F-F'') as well as the analysis of *pnr*-expressing clones in adult heads (Supplementary Figure 20G) confirmed that *pnr* is indeed expressed in the future occipital region.

In summary, the upregulation of *pnr* expression in the developing eye-antennal disc led to larger eyes due to a higher number of ommatidia and a smaller dorsal head cuticle. Therefore, we were able to phenocopy aspects of the "*D. mauritiana*"-like head shape and eye size.

4.4. Discussion

While the genetic architecture of variation in complex morphological traits is being revealed these days (Arif et al., 2013a; Boell and Tautz, 2011; Gaspar et al., 2019; Norry and Gomez, 2017; Pallares et al., 2014; Ramaekers et al., 2018), a mechanistic understanding of how genetic variation affects trait evolution remains largely elusive to date. Here we addressed this gap by combining thorough quantitative phenotyping with comparative transcriptomics, GRN reconstruction and functional genetics to study natural interspecific variation in head shape and eye size between the two closely related *Drosophila* species *D. melanogaster* and *D. mauritiana*.

4.4.1. A developmental model for natural variation in head shape and eye size

Comparative morphology studies revealed that natural intra- and interspecific variation in head shape and eye size is pervasive among species of the *D. melanogaster* subgroup (Gaspar et al., 2019; Hilbrant et al., 2014; Norry et al., 2000; Posnien et al., 2012; Ramaekers et al., 2018). Previous shape analyses suggested that the eyes of *D. mauritiana* are predominantly larger in the dorsal region when compared to *D. melanogaster* (Posnien et al., 2012). Therefore, we restricted our geometric morphometrics analysis to the dorsal head region and found significant natural variation in dorsal head shape. We could confirm that increased eye size in *D. mauritiana* is due to a higher number of ommatidia and goes hand in hand with a reduction of the dorsal interstitial cuticle and a convex bending of the occipital head region. This trade-off between eye size and head cuticle seems to be a common feature of *Drosophila* (Keesey et al., 2019; Norry et al., 2000). Previous attempts to disentangle the genetic architecture of eye and head cuticle size variation did not yet converge on a clear idea, whether the evolution of both structures is linked or not.

Morphological differences in adult traits are a result of variation in developmental processes (Carroll, 2005; Raff, 2000). Since GRNs that regulate such processes are extensively wired, the impact of variation in one node can be elucidated by extensive variation in gene expression (Thompson et al., 2015). Therefore, we applied comparative RNA-seq to reveal 'flexible nodes' in the GRN underlying head and eye development. In accordance with the

previous observation of highly dynamic gene expression throughout eye-antennal disc development (Torres-Oliva et al., 2018), our comparative transcriptomics approach revealed stage-specific interspecific expression divergence. Intriguingly, many of the differentially expressed genes were enriched for binding sites of the GATA transcription factor Pnr that has previously been shown to be involved in dorsal head development (Maurel-Zaffran and Treisman, 2000; Oros et al., 2010). Our finding that *pnr* expression was higher in *D. mauritiana* suggests that natural variation in *pnr* expression may cause extensive remodelling of the transcriptional landscape downstream of this transcription factor.

We could establish a functional link between enhanced *pnr* expression and morphological differences, because overexpression in the dorsal eye-antennal disc of *D. melanogaster* phenocopied major aspects of *D. mauritiana* head shape and eye size. In particular, we observed an enlargement of the dorsal eye area due to increased ommatidia number as well as a reduction of the dorsal interstitial cuticle. Additionally, overexpression of *pnr* resulted in typical convex bending of the occipital region. In contrast, knockdown of *pnr* resulted in the opposite phenotype, characterized by reduction of the eye area, increase of the interstitial cuticle size and massive enlargement of the occipital region. The fact that the strength of the phenotype depended on the RNAi line used, strongly suggests that indeed quantitative differences in *pnr* expression seem to be relevant for phenotypic variation.

Our phenocopy experiment suggests that Pnr is involved in specifying the ratio between retinal tissue and head cuticle. Indeed, at least two major roles of Pnr during *D. melanogaster* eye-antennal disc development have been established. From the late second instar stage on, Pnr regulates the ratio of retinal cells vs. head cuticle cells by suppression of the eye fate in the dorsal region of the eye-antennal disc. This suppression is either accomplished by directly repressing members of retinal determination network as for instance *teashirt (tsh)* or indirectly via activation of *wingless (wg)* (Oros et al., 2010). Our results combining transcriptomics, ATAC-seq and transcription factor binding motif enrichment did not identify *tsh* as a direct target gene of Pnr, suggesting that the observed repression of *tsh* by Pnr (Oros et al., 2010) may be indirect. Interestingly, with *eyeless* (*ey*) and *eyegone* (*eyg*) we found two other members of the retinal determination network among the putative Pnr targets. Whether potential direct interactions are negative and may be linked to the repression of retinal fate in the dorsal disc remains to be determined. This later role in defining the dorsal cuticle *vs.* retinal fate is well in

line with the observed trade-off between eye size and interstitial cuticle. Additionally, our tracing experiment revealed that *pnr*-expressing cells contribute to the dorsal occipital head region. Therefore, a direct effect on the morphological differences in this region is likely.

During early eye-antennal disc development, Pnr plays a pivotal role in defining the dorsal/ventral boundary and is therefore responsible for overall tissue growth (Maurel-Zaffran and Treisman, 2000; Singh et al., 2005; Singh and Choi, 2003). Our result that *ey* is among the putative direct Pnr target genes offers now an exciting and yet unpredicted early role of Pnr in *ey* activation in the peripodial epithelium and in margin cells. It has recently been shown that Ey activity in the peripodial epithelium and the margin cells is necessary for *decapentaplegic* (*dpp*) induction and subsequent initiation of the morphogenetic furrow (Baker et al., 2018). Loss of Ey function also interferes with the placement of the dorsal/ventral boundary (Baker et al., 2018) providing a functional link to this well-established early role of Pnr. Throughout the third larval instar Pnr is predominantly expressed in the peripodial epithelium and our lineage tracing experiment showed that during earlier stages *pnr* must be expressed in cells that contribute to the dorsal posterior margin where the morphogenetic furrow is initiated. Therefore, Pnr is expressed in the right cells at the right time to act upstream of *ey* during dorsal/ventral boundary establishment and the initiation of the morphogenetic furrow, suggesting that differences in early *pnr* expression could have a direct effect on retinal development.

In summary, we provide a comprehensive developmental model suggesting that variation in expression of a pleiotropic central transcription factor is responsible for the concerted diversification of a complex morphological trait.

4.4.2. Pnr and Ush represent a functionally linked pleiotropic module in the GRN underlying head and eye development

Our developmental data showed that natural variation in *pnr* expression influences different developmental processes. Our combinatorial RNA-seq and ATAC-seq data revealed that more than 1,000 putative Pnr target genes expressed during eye-antennal disc development, further substantiating its central role during head development. Some of the target genes showed expression profiles in agreement with an activating role of Pnr, while some targets showed signatures of a negative relationship. This observation suggests that the dual regulatory role of Pnr observed in the wing disc (Fromental-Ramain et al., 2010, 2008) may be

true for the eye-antennal disc as well. The repressive role of Pnr in the wing imaginal disc is realized upon heterodimerization with its co-factor Ush (Cubadda et al., 1997; García-García et al., 1999; Haenlin et al., 1997). However, it was thought that *ush* was not expressed in the developing eye-antennal disc (Fossett et al., 2001) or non-functional (Maurel-Zaffran and Treisman, 2000). Following this assumption, *ush* overexpression was in fact mainly used to mimic *pnr* knock-down (Fossett et al., 2001). Based on our RNA-seq data we show for the first time that *ush* is transcribed in the eye-antennal disc. Additionally, we confirm that *ush* transcripts are translated and that the protein is co-localized with Pnr in the squamous cells of the dorsal peripodial epithelium and in the cuboidal cells of the disc margin. Furthermore, *ush* expression is necessary for proper head development, since knock down in the dorsal part of the eye-antennal disc resulted in irregularities in adult dorsal head cuticle and head bristle pattern. The latter effect has been previously described for *ush* hypomorphs (Cubadda et al., 1997). Intriguingly, overexpression of Ush in the dorsal eye-antennal disc resulted in a double antenna phenotype reminiscent of that observed upon loss of Pnr function (Oros et al., 2010).

The co-expression as well as similar functions of *pnr* and *ush* strongly suggest that they interact during eye-antennal disc development. This hypothesis is further supported by a clear genetic interaction between both factors. We showed that Pnr is involved in *ush* activation. Since we did not find *ush* as a potential target gene of Pnr, the activation of *ush* may be indirect. Furthermore, we identified an autoregulatory loop of Pnr that seems to be negatively modulated by the presence of Ush. Since we found *pnr* in our list of putative Pnr target genes, we propose here that the expression level of *pnr* is kept in balance via activation by Pnr alone and repression by the Pnr-Ush heterodimer.

This model suggests that the various roles of Pnr during eye-antennal disc development could be facilitated by the stoichiometry between Pnr and its co-factor Ush. For instance, the early function of Pnr in dorsal/ventral boundary establishment and morphogenetic furrow initiation is most likely independent of Ush (i.e. mainly activating role of Pnr). This is supported by our observation that reduction of *pnr* expression via RNAi did not influence the final ommatidia number in the adult eyes. In the absence of Ush the reduced *pnr* expression can be compensated by an increased auto-activation to restore normal retinal development. Additionally, the effect of *ush* RNAi was mostly restricted to the dorsal head cuticle, suggesting that it might not play a major role during retina development. However, the later function in

head cuticle development and sensory bristle formation most likely depends on the ratio of Pnr and Ush. This is supported by a similar expression profile of *pnr* and *ush* during third instar development. Additionally, it has been shown that sensory bristles in the thorax arise at regions with high *pnr* and low *ush* expression (Cubadda et al., 1997; Heitzler et al., 1996). Our overexpression of Pnr using the *VT042374* driver line consistently resulted in duplication of the posterior vertical bristles, underpinning the role of Pnr in sensory bristle formation (Heitzler et al., 1996; Ramain et al., 1993). Interestingly, this is reminiscent of the phenotype described for a dominant pnr^{D} allele (Heitzler et al., 1996), which is characterized by a loss of the ability to dimerize with Ush. Since Ush antagonizes bristle formation (Haenlin et al., 1997), the duplication of the posterior vertex bristles is most likely the result of overexpression in the posterior part of the dorsal peripodial epithelium where endogenous *ush* is not expressed anymore. In the anterior region, the endogenous *ush* expression is sufficient to block the development of additional sensory bristles. In contrast, overexpression of *ush* in most of the dorsal peripodial epithelium did not only result in the loss of the posterior vertical bristles, but also in the loss of the anterior vertical bristles, suggesting that extra Ush above a certain threshold completely antagonizes sensory bristle formation. Hence, the correct stoichiometry between Pnr and its co-factor Ush is crucial for proper dorsal head and sensory bristle formation. This notion is further supported by our observation that also *ush* to be slightly upregulated in *D. mauritiana*, recapitulating the expression dynamics of *pnr*.

In summary, we identified variation in expression of a highly pleiotropic regulatory module composed of Pnr and Ush that causes the differential expression of a plethora of potential target genes. Therefore, we conclude that this regulatory module might be a flexible node in the GRN underlying head and eye development in *Drosophila*.

4.4.3. GRN rewiring facilitates natural variation in pleiotropic developmental factors

Eye-antennal disc development is highly complex and the underlying GRN is extensively rewired both throughout time (Torres-Oliva et al., 2018) and in different parts of the disc (Potier et al., 2014). For instance, genes of the retinal determination network are required for the initial proliferation and growth of the entire eye-antennal disc and later they play a pivotal role in retinal specification (Baker et al., 2018; Bessa, 2002; Lopes and Casares, 2010; Neto et al., 2017). It has been suggested that the retinal determination genes are part of different GRNs during these events and extensive rewiring of the GRNs allows them to fulfil temporally

restricted tasks (Palliyil et al., 2018). Similarly, the integration of gene products in spatially restricted GRNs may also explain why some genes are broadly expressed in the eye-antennal disc although they regulate different processes in different parts of the disc, which give rise to the various head structures (Palliyil et al., 2018; Potier et al., 2014). It seems therefore that rewiring of GRNs facilitates the use of the same developmental gene products in different contexts.

The various described roles for Pnr (summarized in (Oros et al., 2010)), its continuous expression in the eye-antennal disc and the observation that variation in *pnr* expression affects overall head shape and eye size simultaneously, strongly suggest that Pnr is involved in several GRNs during eye and head development. The interaction with co-factors, such as Ush provides a mechanism facilitating such network rewiring by modulating the role of Pnr from an activating to a repressing transcription factor. We conclude that the dynamic nature of GRNs may explain how interspecific variation in expression of a highly pleiotropic and central transcription factor such as Pnr can result in extensive remodelling of the transcriptomic landscape in an otherwise tightly controlled GRN.

Intriguingly, Pnr is not the only central and pleiotropic factor implicated in natural variation in head shape and eye size. It has recently been shown that a single point mutation in the *cis*-regulatory sequence of *ey*, one of the most upstream factors of the retinal determination network (Callaerts et al., 1997), leads to heterochronic changes in its regulation and subsequent variation in eye size among *D. melanogaster* laboratory strains. This polymorphism segregates in natural *D. melanogaster* populations and it shows signatures of longitudinal cline in Europe, suggesting that it may provide some selective advantage in certain environmental conditions (Ramaekers et al., 2018). In summary, we hypothesize that the modularity of regulatory interactions during development may allow selection to act on highly pleiotropic developmental factors to drive diversification of complex morphological traits.

4.4.4. Evolution of GRNs and implications for convergent evolution of head shape and eye size

A trade-off between the size of the compound eye and other head structures is common in *Drosophila* (Keesey et al., 2019; Norry et al., 2000). Depending on the environment, enlargement or reduction of the eye is most probably selected, since smaller eyes and less ommatidia lead indeed to poorer temporal acuity (Currea et al., 2018; Ramaekers et al., 2018)

and has ecological implications (Currea et al., 2018). This assumption is also supported by the fact that an enlargement of the compound eye is associated with increased optic lobe size (Keesey et al., 2019). However, functional sensory systems consume tremendous amounts of energy (Niven and Laughlin, 2008; Tan et al., 2005) suggesting that their size must be tightly controlled. It has been proposed that the common origin of the adult visual (i.e. compound eyes) and olfactory (i.e. antennae) system from the same imaginal disc provides an opportunity to balance the energy investment either in olfactory or in visual structures (Keesey et al., 2019). Although *D. mauritiana* was not included in this large-scale survey, it is likely that the resource allocation hypothesis applies to this species as well. However, it remains to be studied how temporal acuity and the size of visual neuropils coevolved with head shape variation between *D. melanogaster* and *D. mauritiana*.

In the light of a pervasive trade-off between eye and head cuticle in *Drosophila* it is tempting to ask whether this morphological trait evolves through the same or different nodes of the underlying GRN among different populations or species. Between *D. melanogaster* and *D. simulans* different QTL regions were identified for eye size and the width of the interstitial cuticle. This observation was supported by quantitative developmental data showing that the anlagen for the head cuticle start to diverge in size prior to the retinal tissue (Arif et al., 2013a). Therefore, the trade-off seems to be regulated by independent factors in these two species. However, recent quantitative genetics analyses identified some loci that affect eye size and head cuticle in opposite directions in intraspecific comparisons in *D. melanogaster* and *D. simulans* (Gaspar et al., 2019; Norry and Gomez, 2017). Additionally, our finding that variation in *pnr* expression influences both traits simultaneously further suggests that they may be genetically linked in *D. melanogaster* and *D. mauritiana*. Therefore, a convergent evolution of the trade-off in *Drosophila* is likely. A detailed analysis of the morphological basis of eye size differences showed that bigger eyes can be the result of differences in ommatidia number (e.g. between *D. melanogaster* and *D. mauritiana*) or ommatidia size (e.g. between *D. simulans* and *D. mauritiana*) (Posnien et al., 2012). Since these two features are regulated at different time-points and developmental processes (reviewed in (Amore and Casares, 2010; Domínguez and Casares, 2005)) it is conceivable that the molecular and developmental basis of eye size differences varies in different groups. In summary, our current knowledge based on quantitative genetics, developmental as well as morphological data suggests that different

nodes within the GRN underlying head and eye development may evolve to give rise to variation in head morphology in *Drosophila.*

4.4.5. Conclusion and Outlook

We provide here a methodological framework to reveal flexible nodes within GRNs and to subsequently validate these findings. Our comparative transcriptomics approach can be used as entry point to study the evolution of complex morphological traits or it can be applied to link already identified genetic variation to nodes within developmental GRNs and to developmental processes. It is important to note, however, that this approach unfolds its full potential if complemented with quantitative genetics data that allows identifying exact genetic variants associated with trait variation. The fact that we were not able to phenocopy the *D. mauritiana* head shape and eye size entirely, suggests that multiple genomic loci are responsible for the observed morphological divergence between *D. melanogaster* and *D. mauritiana*. Furthermore, it remains to be established, whether the *pnr* and/or *ush* loci contain genetic variants associated with eye size and head shape differences. Quantitative genetics approaches are not applicable since interspecific crosses between *D. melanogaster* and *D. mauritiana* result in infertile F1 females. However, reciprocal hemizygosity tests (Stern, 2014) for Pnr, Ush and putative regulators of these two factors represent a powerful tool to further dissect the causative genetic variants in the future. Overall, much more genetic as well as developmental data from different groups is necessary to draw a full picture of this exciting morphological phenomenon. Eventually, it remains to be established, whether similar functional requirements and ecological forces are involved in shaping the *Drosophila* head morphology.

4.5. Material and Methods

4.5.1. Generation of the transcriptomic dataset

Flies from the following strains were raised at 25°C at a 12:12 dark:light cycle for at least two generations and their eggs were collected on agar plates for one hour: *D. melanogaster* (OreR), *D. mauritiana* (TAM16). 30 L1 larva were collected in vials and developing eye-antennal discs were dissected at 72h AEL (120–130 discs; m and f), 96h AEL (80–90 discs; f) or 120h AEL (40-50 discs; f) and stored in RNALater (Quiage, Venlo, Netherlands). For each species and stage 3 biological replicates were generated. Total RNA was isolated using the Trizol (Invitrogen, Thermo Fisher Scientific, Waltham, Massachusetts, USA) method according to the manufacturer's recommendations and the

samples were DNAseI (Sigma, St. Louis, Missouri, USA) treated in order to remove DNA contamination. RNA quality was determined using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) microfluidic electrophoresis. Only samples with comparable RNA integrity numbers were selected for sequencing. Library preparation for RNA-seq was performed using the TruSeq RNA Sample Preparation Kit (Illumina, catalog ID RS-122-2002) starting from 500 ng of total RNA. Accurate quantification of cDNA libraries was performed using the QuantiFluordsDNA System (Promega, Madison, Wisconsin, USA). The size range of final cDNA libraries was determined, applying the DNA 1000 chip on the Bioanalyzer 2100 from Agilent (280 bp). cDNA libraries were amplified and sequenced (50 bp single-end reads) using cBot and HiSeq 2000 (Illumina). Sequence images were transformed to bcl files using the software BaseCaller (Illumina). The bcl files were demultiplexed to fastq files with CASAVA (version 1.8.2)

4.5.1.1. Mapping

The reads were mapped against strain-specific transcriptomes of *D. melanogaster* and *D. mauritiana*, including CDS and UTRs (Torres-Oliva et al., 2016) using Bowtie2 v. 2.3.4.1 with the following parameters: `-very-sensitive-local -N1` (Langmead et al., 2009). Samtools version 1.9 was used to further process the reads, and count the reads mapped to each transcript (`idxstats`) (Li et al., 2009).

4.5.1.2. DEA and data visualization

A principal component analysis was done using the regularized log (rlog) transformation from the DESeq2 package (DESeq2_1.22.2; R version 3.5.2) (Love et al., 2014).

We then used DeSeq2 (Love et al., 2014) to perform a pairwise differential expression analysis between the two species at each time point (*D. melanogaster* 72h *vs. D. mauritiana* 72h, *D. melanogaster* 96h *vs. D. mauritiana* 96h, *D. melanogaster* 120h *vs. D. mauritiana* 120h). We used the online tool Metascape (Zhou et al., 2019) to perform GO enrichment analysis for each time point. All genes that were significantly differentially expressed (log2FC > 0 | log2FC < 0 and padj < e0.05) in at least one stage (8350 unique genes) were combined and clustered using the coseq package (version 1.6.1; (Rau et al., 2013)) with the following parameters: `K=2:25, transformation="arcsin", norm="TMM", model="Normal"`. We searched for potential upstream factors in 5kb upstream of the TSS, 5'UTR regions and 1st introns using the i-cisTarget tool (Imrichová et al., 2015; Pereira et al., 2006) keeping the default parameters:

Minimum fraction of overlap: 0.4., NES: 3.0, ROC threshold for AUC calculation: 0.01. Metascape was used to analyse differential enrichment of GO terms for each pairwise comparison.

4.5.2. Generation of the ATAC-seq dataset

For the generation of ATAC-seq datasets we followed (Buenrostro et al., 2013). Developing eye-antennal discs of *D. melanogaster* were dissected in ice-cold PBS at 72h, 96h and 120h AEL. PBS was removed and exchanged for 50 µl lysis buffer (10 mM Tris-HCl (pH = 7.4); 10 mM NaCl; 3 mM MgCl2; 0.1 % IGEPAL). The mixture was pipetted several times up and down to lyse the cells and then split into micro centrifuge tubes. Centrifugation for 10 min at 500 g and 4 °C. The cell number was assessed in one of the samples and between 50,000 and 80,000 nuclei were used in subsequent steps. The supernatant was removed and the pellet(s) dissolved in 47.5 µl 1X tagmentation buffer (20 mM Tris-CH3COOH (pH = 7.6); 10 mM MgCl2; 20 % (vol/vol) dimethylformamide) with 2.5 µl Tn5 Transposase and then incubated for 30 min at 37 °C. For purification we used the QIAGEN MinElute Kit and eluted in 10 µl Elution Buffer (10 mM Tris, pH = 8). For the PCR amplification was done as follows:

- 10 µl tagmented chromatin
- 10 µl H2O
- 2.5 µl Nextera PCR primer 1*
- 2.5 µl Nextera PCR primer 2**
- 25 µl NEBNext High-Fidelity 2X PCR Master Mix (Cat #M0541)

We used the following program:

(1) 72 °C 5 min
(2) 98 °C 30 sec
(3) 98 °C 10 sec
(4) 63 °C 30 sec
(5) 72 °C 1 min
(6) repeat 3-5 13 times
(7) hold at 4 °C

followed by another 2x purification step with the QIAGEN MinElute Kit: elution in 2 X 10 µl Elution Buffer (10 mM Tris, pH = 8).

* AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG

** Ad2.2_CGTACTAG CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
Ad2.3_AGGCAGAA CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4_TCCTGAGC CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5_GGACTCCT CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
Ad2.6_TAGGCATG CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
Ad2.7_CTCTCTAC CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT

4.5.3. Bioinformatics processes of the ATAC-seq dataset

We performed quality checks of the sequenced reads using FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The reads were trimmed, using Trimmomatic (version 0.36) (Bolger et al., 2014) appyling a sliding window trimming with the parameters `slidingwindow 4:15` and `minlen 30`. Trimmed reads were mapped to the *D. melanogaster* genome (version 6.13) after discarding the mitochondrial genome, using Bowtie2 (version 2.3.4.3) (Langmead et al., 2009), with the commands: `--no-unal` and `-X2000`. Samtools version 1.9 (Li et al., 2009) were subsequently used to convert the sam to bam files, and to sort and index bam files. We removed duplicates using PICARD (version 2.1.1, https://github.com/broadinstitute/picard) with default parameters and converted the resulted bam files to bed files. Reads were then centered as described in (Buenrostro et al., 2013). We used MACS2 (version 2.1.2) (Zhang et al., 2008, p. 2) with the following commands `-g dm --nomodel --shift -100 --extsize 200 -q 0.01 –bdg` to call significant peaks. We used the Integrated Genome Browser (IGB, (Freese et al., 2016)) to visualize the read depth and peaks. Peaks were annotated to the closest gene using the `annotatePeaks.pl` program from the HOMER software package (v4.8.3) using dm6 as genomic input.

4.5.4. Definition of a Pnr target gene list

As a basis for the high confidence list of putative Pnr target genes a Chip-chip dataset was used (downloaded on 1st of July, 2015 from http://furlonglab.embl.de/data/download, (Junion et al. 2012), which comprises ChIP-chip experiments in the *Drosophila* embryo with several transcription factors, including Pnr at two time points (4-6h AEL and 6-8h AEL). All Pnr-binding regions from both time points were selected with a Tile-Map score of <5.5. and where the distance of the center of the peak to the TSS was -1000 bp and +1000 bp. This resulted in a gene list of 4009 putative Pnr targets (Figure 10). The peak regions of these genes were used to search for the Pnr GATA motif, resulting in a list of 1675 putative target genes. We restricted the list of potential Pnr target genes to those genes which are expressed (>= 10 reads on average for each stage in *D. melanogaster*) in our transcriptomic dataset and

performed hierarchical clustering using coseq (Rau and Maugis-Rabusseau, 2017) according to their expression dynamics with the following parameters: `K=2:25, transformation="arcsin", norm="TMM", model="kmeans"`. We downloaded a list of all known direct (TF-gene) or genetic interactions of Pnr from the DroID database (Yu et al., 2008) and found an overlap of 30 genes, of which 21 are differentially expressed between the *D. melanogaster* and *D. mauritiana*. We used Cytoscape (Shannon et al., 2003) to visualize the interaction between these target genes and potential upstream regulators found in the database.

4.5.5. qPCR

Discs from *D. melanogaster* and *D. mauritiana* larva were dissected in ice-cold PBS at 96h AEL and 120h AEL and collected in TRIZOL. The samples were then homogenized using a TissueLyser and total mRNA was extracted using the Phenol/Chloroform extraction method. We then used TurboDNAse to remove potential gDNA contamination. Concentration was measured using Nanodrop and the Maxima First Strand cDNA Synthesis Kit for RT-qPCR kit was used for cDNA preparation. To test the efficiency of primers, we prepared four 1:4 dilutions of a pool of all RNA samples per species (1:2, 1:6, 1:18, 1:54; for calculations see Supplementary Figure 11). Real-time qPCR was then performed using the HOT FIREpol EvaGreen qPCR Mix Plus (ROX) (Solis BioDyne, Tartu, Estland) and a CFX96 Real-time PCR System (Bio-Rad Laboratories, Hercules, CA, USA). Log2 fold changes in expression were calculated using the ΔΔCT method (Livak and Schmittgen, 2001) with actin79b as a reference gene.

Following primer pairs were used:

pnrB: f: CGCAGACGAATCAAACG, r: TCACGTTCTGATAGTCGC

actin 79b: f: CGCAAGGATCTGTATGCCAAC, r: TCTTGATGGTGGACGGGG

The following temperatures were used:

1) 95°C – 15:00
2) 95°C – 00:30
3) 56°C – 0:30
4) 72°C – 0:30
5) Repeat step 2-4 for 39x
6) 65°C – 0:05

7) 95°C

4.5.6. Antibody staining

Developing eye-antennal discs were dissected in ice-cold PBS and fixed for 30 min in 4% paraformaldehyde (PFA). The discs were then washed 3 times in 0.03% PBT (Phosphate buffered saline 1%, Triton X-100) before blocked in 5% goat serum for 30min. Incubation with the Primary Antibody was done for 90 min, before 3 additional washing steps with PBT and one round of blocking for 30 min. The tissue was then incubated overnight with the Secondary Antibody on a rocking plate on 4°C. If needed, Phalloidin-488 (1:100) was added. After 3 washing steps with PBT, the discs were incubated with DAPI (1:1000) for 10 min, followed by one washing step with PBT and one washing step with PBS. Subsequently, the discs were mounted in mounting medium (80% glycerol + 4% n-propyl-galate) and kept at least one night on 4° degree before imaging.

Antibodies used: We generated polyclonal, primary antibodies against Pnr (Junion et al., 2012) (Pnr_B (125-294) (TPLWRRDGTGHYLCNACGLYHKMNGMNRPLIKPSKRLVSATATRRMGLCCTNCGTRTTTLWRRNNDGEPVCNACGLYYKLHGVNRPLAMRKDGIQTRKRKPKKTGSGSAVGAGTGSGTGSTLEAIKECKEEHDLKPSLSLERHSLSKLHTDMKSGTSSSSTLMGHHSAQQ) and Pnr_B (206-336) (GVNRPLAMRKDGIQTRKRKPKKTGSGSAVGAGTGSGTGSTLEAIKECKEEHDLKPSLSLERHSLSKLHTDMKSGTSSSSTLMGHHSAQQQQQQQQQQQQQQQQQQQQQSAHQQCFPLYGQTTTQQQHQQHGH)) and Ush (Fossett et al., 2001) based on previous knowledge, against the peptide sequences: Ush-(231–250) (CSHRIKDTDEAGSDKSGAGG) and Ush-(1174 –1191) (VGGHGQQKNKENLQEAAI). Before usage, both antibodies were preabsorbed overnight on *Drosophila* embryos on 4°C.

Please note that we confirmed the specificity of the Ush antibody by recapitulating known Ush expression domains in the wing imaginal disc (Supplementary Material Figure 1A) and during embryonic development (Supplementary Material Figure 1B) (Muratoglu et al., 2006; Tomoyasu et al., 2000)). For test stainings in embryos we collected embryos for several hours on apple agar plates, removed the chorion with 50% Klorix and rinsed them 3x with 0.03% PBT (Phosphate buffered saline 1%, Triton X-100). We fixed the embryos with heptane and 2% formaldehyde for 20min and washed with MeOH, followed by washing steps with PBT. The embryos were then blocked in 3% BSA for one hour, followed by incubation with the primary AB overnight. After two washing steps with PBT, we added HRP-coupled secondary AB for 90

min. After three washing steps with PBT we performed a DAB (3'-3diaminobenzidine) staining. The embryos were then washed again 2 times in PBT and mounted in glycerol.

Concentrations Primary AB's: Anti-Pnr (rabbit): 1:200, Anti-Ush (rabbit): 1:2000, Anti-GFP (chicken): 1:1000; Concentrations Secondary AB's: Anti-Chicken-488; Anti-Rabbit-Cy3: 1:500; Anti-rabbit-HRP-coupled (1:1000).

Pictures of eye-antennal discs upon antibody staining were taken using a Zeiss LSM 710 confocal microscope. Antibody stainings were visualized and processed with Fiji software (Schindelin et al., 2012). Vertical section of the confocal pictures were generated using the Volume Viewer plugin (https://imagej.nih.gov/ij/plugins/volume-viewer.html) with the following parameters: Display Mode: `Slice and Boarders`, Interpolation: `Nearest Neighbour`, Transfer Function: `Fire LUT`.

4.5.7. Geometric Morphometrics

We imaged WT species, each parental line and the offspring of the respective crosses from the dorsal view of the head using a Leica M205 FA stereo microscope. We placed 64 landmarks on pictures of these dorsal heads using the tpsDig2 software. We then defined 23 fixed landmarks and 41 sliding landmarks using tpsUtil. tpsRelw32 was used to calculate the consensus (i.e. Procrustes superimposition), partial warps and relative warps (https://life.bio.sunysb.edu/morph/). Using MorphoJ (Klingenberg, 2011) for visualization, we performed Procrustes Fit and generated a covariance matrix. To analyse differences in dorsal head shapes in WT *D. melanogaster* and *D. mauritiana*, we performed discriminate function analysis using MorphoJ. We further performed a principal component analysis (PCA) to analyse difference in head shapes upon knock-down or up-regulation of *pnr*.

4.5.8. Overexpression/Knock-down of *pnr* and *ush*

To overexpress or knock-down *pnr* and *ush*, the following fly lines were used:

D. melanogaster (Oregon R) and *D. mauritiana* (TAM 16) (both kindly provided by Prof. Alistair McGregor, Oxford Brookes University), *pnr* GAL4/TM6B (kindly provided by Prof. Marc Haenlin), *pnr* GAL4>UAS2YFP/TM6B (kindly provided by Prof. Marc Haenlin, CBI Toulouse), *ush* GAL4 26662 (y[1] w[*]; P{w[+mW.hs]=GawB}ush[MD751]), Bloomington; UAS *ush* 14IIA/CyO (kindly provided by Prof. Marc Haenlin, CBI Toulouse); ush-RNAi 3622 (y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.HM05193}attP2/TM3, Sb[1]), Bloomington;UAS *pnr* (w; UAS-*pnr*/CyO;

TM2/TM6B) (kindly provided by Prof. Fernando Casares); *pnr*-RNAi (VT101522/KK, #108962 VDRC Stock Center and VT6224/GD, #1511 VDRC Stock Center); oc2-Gal4/CyO (kindly provided by Prof. Fernando Casares).

All crosses were performed at 25°C and at a constant 12h/12h light/dark cycle. Since Pnr and Ush are crucial during embryonic development, we chose combinations of GAL4/UAS lines, that resulted in a phenotype but were not lethal during embryonic, larval of pupal stages. We used a set of GAL4 lines, that overlap a range of weak to strong driving capacity and in the case of *pnrRNAi* we used two RNAi lines with different effector strengths.

4.5.9. *pnr* expression and lineage.

The *pnr-GAL4* line *pnrMD237* (Calleja et al., 1996), recombined with UAS-GFP, was used to follow *pnr* expression in imaginal and pupal eye-antennal discs. Adult *pnr* expression domain in adult heads was monitored in *y; pnr-GAL4/UAS-y+* flies (Calleja et al., 1996) as the cuticle region with y-rescued pigmentation. To follow the *pnr-GAL4* lineage, *pnr-GA4*, UAS-GFP flies were crossed to UAS-flipase; *act5c>stop< nuc-lacZ* flies (Struhl and Basler, 1993). In the discs of the progeny, *pnr* expression was visualized with anti-GFP and its lineage with anti β-galactosidase.

4.5.10. Immunostaining and imaging.

Third instar or pupal discs were processed as in (Magri et al., 2018). Primary antibodies were chicken anti-GFP (1/500; ab13970, Abcam), rabbit anti-β-galactosidase (1/1000; Cappel) and mouse anti-Eya (1/500; 10H6, Developmental Studies Hybridoma Bank, Iowa University). Secondary antibodies at 1/400 were from Molecular Probes. Imaging was carried out on a Leica SPE confocal setup (ALMI, CABD).

4.5.11. Adult head cuticle preparation.

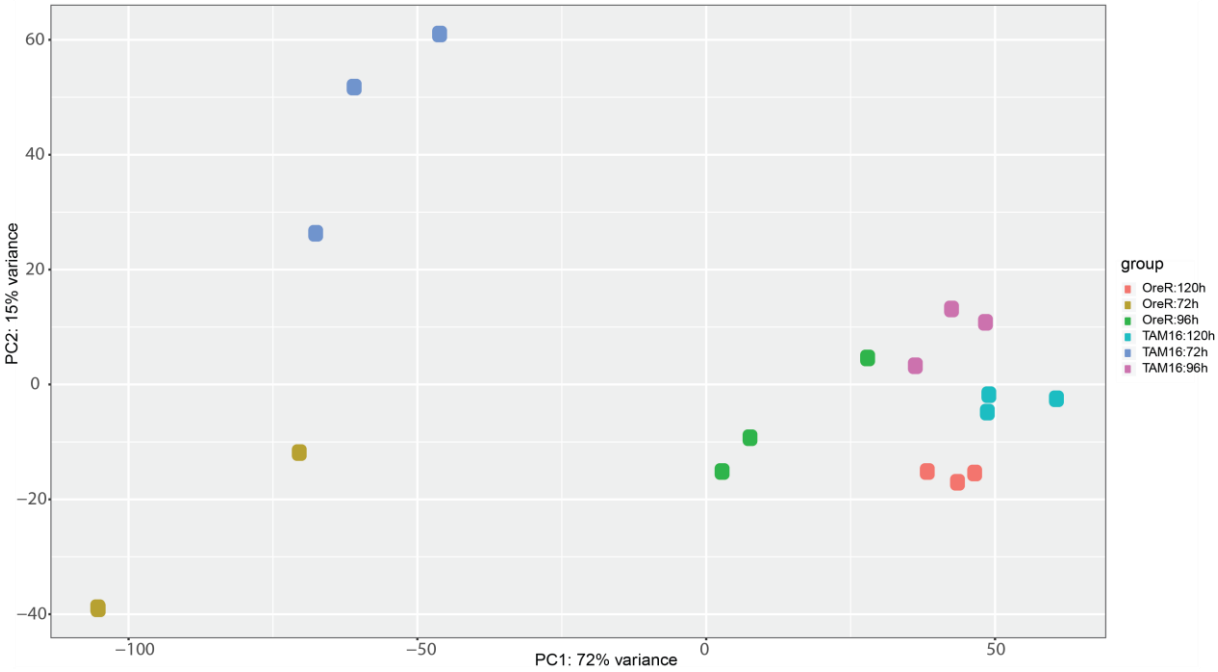
Dissected adult heads in PBS were mounted in Hoyer's mountant: Lactic Acid (50:50) as in (Magri et al., 2018).

4.5.12. Ommatidia Counting

To estimate ommatidia number of single fly eyes, we took pictures of one eye per fly (50 stacks/eye) using a Leica M205 FA stereo microscope and an external light source, which resulted in reflection of light by each ommatidium. We used FIJI to perform the following analyses. We performed Z-projection using maximum intensity and then transformed each picture, so that the single reflection of each ommatidium is represented by a black dot. We

then used the ICTN cell counter tool (with the following parameters: Width: 7, Minimum Distance: 17, Threshold: 1.5) to estimate the number of black dots, i.e. the number ommatidia. Statistical analysis of ommatidia number was done using a one-way ANOVA and pair-wise comparisons were calculated using Tukey HSD test.

4.6. Supplementary Figures

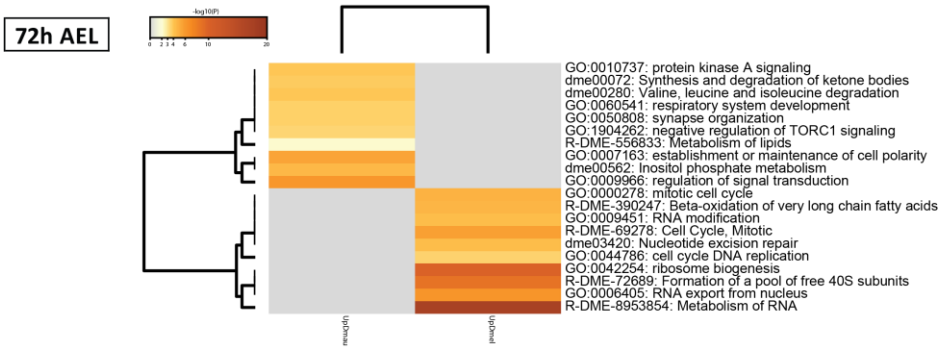


Supplementary Figure 8. Principal component analysis of all RNA-seq samples based on rlog transformed read counts. PC1 separates the samples according to time-points, whereas PC2 separates the data mainly by species. *D. melanogaster* (OreR), *D. mauritiana* (TAM16).

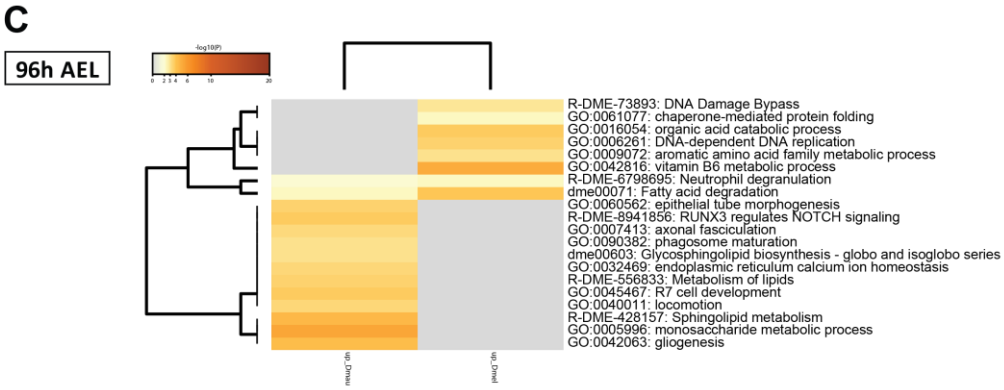
A

	upregulated in *D. melanogaster*	upregulated in *D. mauritiana*	total
72h AEL	3244	3439	6683
96h AEL	1622	1638	3260
120h AEL	1130	1250	2380

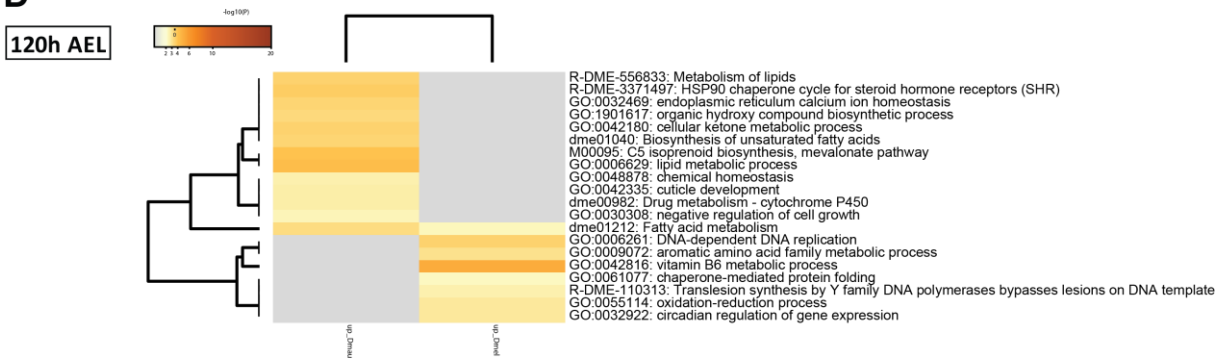
B



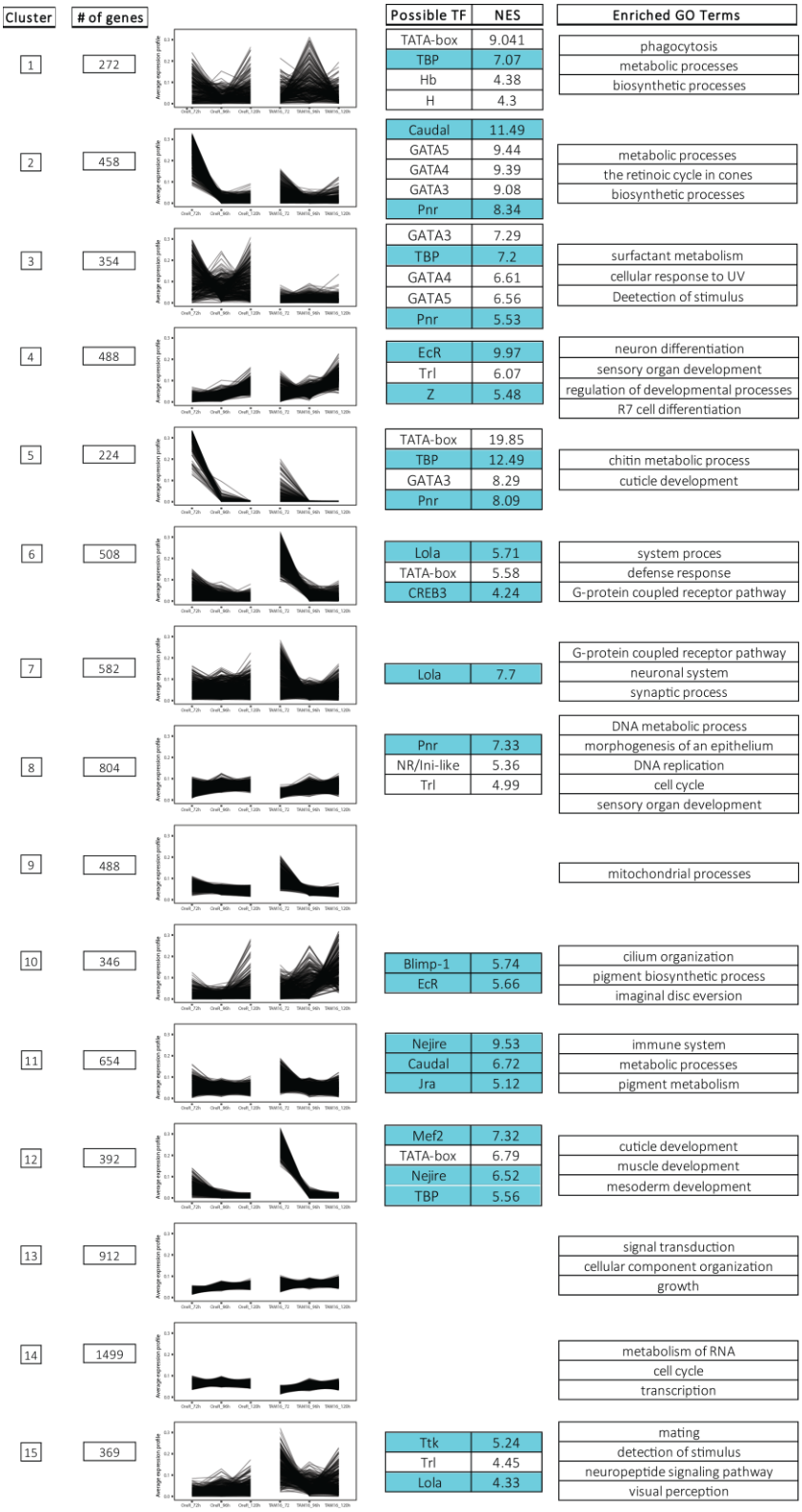
C



D



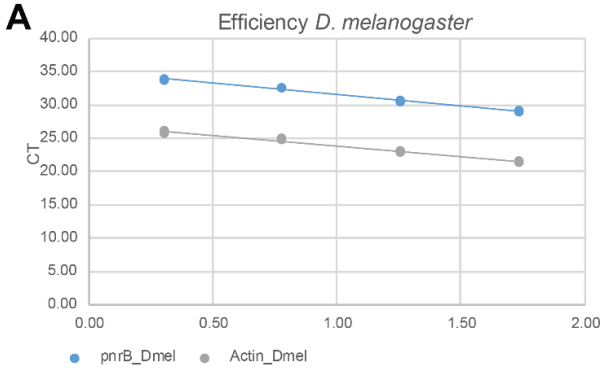
Supplementary Figure 9. A. Number of genes that are significantly differentially expressed (padj <= 0.05) between the two species for each time-point. *D. melanogaster* is depicted in red and *D. mauritiana* in blue. Total number of differentially expressed gene that are upregulated are shown in the third column, as a sum of genes, upregulated in each species. **B.** Differential GO-term analysis of genes that are upregulated in each species per time-point.



Cluster	# of genes	Possible TF	NES	Enriched GO Terms
1	272	TATA-box	9.041	phagocytosis
		TBP	7.07	metabolic processes
		Hb	4.38	biosynthetic processes
		H	4.3	
2	458	Caudal	11.49	metabolic processes
		GATA5	9.44	the retinoic cycle in cones
		GATA4	9.39	biosynthetic processes
		GATA3	9.08	
		Pnr	8.34	
3	354	GATA3	7.29	surfactant metabolism
		TBP	7.2	cellular response to UV
		GATA4	6.61	Deetection of stimulus
		GATA5	6.56	
		Pnr	5.53	
4	488	EcR	9.97	neuron differentiation
		Trl	6.07	sensory organ development
		Z	5.48	regulation of developmental processes
				R7 cell differentiation
5	224	TATA-box	19.85	chitin metabolic process
		TBP	12.49	cuticle development
		GATA3	8.29	
		Pnr	8.09	
6	508	Lola	5.71	system proces
		TATA-box	5.58	defense response
		CREB3	4.24	G-protein coupled receptor pathway
7	582	Lola	7.7	G-protein coupled receptor pathway
				neuronal system
				synaptic process
8	804	Pnr	7.33	DNA metabolic process
		NR/Ini-like	5.36	morphogenesis of an epithelium
		Trl	4.99	DNA replication
				cell cycle
				sensory organ development
9	488			mitochondrial processes
10	346	Blimp-1	5.74	cilium organization
		EcR	5.66	pigment biosynthetic process
				imaginal disc eversion
11	654	Nejire	9.53	immune system
		Caudal	6.72	metabolic processes
		Jra	5.12	pigment metabolism
12	392	Mef2	7.32	cuticle development
		TATA-box	6.79	muscle development
		Nejire	6.52	mesoderm development
		TBP	5.56	
13	912			signal transduction
				cellular component organization
				growth
14	1499			metabolism of RNA
				cell cycle
				transcription
15	369	Ttk	5.24	mating
		Trl	4.45	detection of stimulus
		Lola	4.33	neuropeptide signaling pathway
				visual perception

Supplementary Figure 10. Clustering of all genes that are significantly differentially expressed (padj <= 0.05) between the two species in at least one time-point resulted in 15 distinct co-expression profiles. Shown is the number of genes in each cluster and the expression profile plot. The tables show transcription factors whose transcription factor binding motifs were enriched 5kb upstream of the TSS, 5'UTR regions and 1st introns of the clustered genes. The NES-factors are shown in the second column. Transcription factors labelled in cyan are themselves significantly differentially expressed in at least one stage. GO-terms enriched in each cluster are given in the last column.

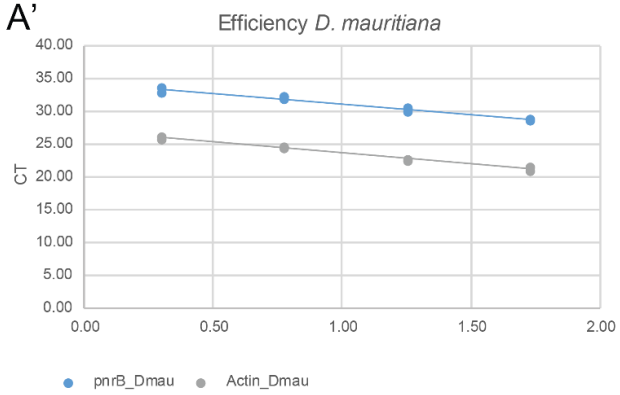
A



y =-3.3722x + 34.946
R² = 0.9896
y =-3.1991x + 27.063
R² = 0.9886

	slope	E[%]
pnrB_Dmel	-3.37	98.03
Actin_Dmel	-3.19	105.82

A'



y =-3.2395x + 34.408
R² = 0.9726
y =-3.3155x + 26.968
R² = 0.9808

	slope	E[%]
pnrB_Dmau	-3.23	103.98
Actin_Dmau	-3.31	100.50

B

	Average_Replicates - 96h					
	a	b	c			
Dmel_pnrB		30.62	30.17			
Dmel_Actin		21.74	21.39	Average_Δct		
Δct_Dmel_pnrB		8.88	8.79	8.83		
Dmau_pnrB	30.45	30.36	29.27			
Dmau_Actin	21.75	21.92	21.16	Average_Δct	ΔΔct	2^-ΔΔct
Δct_Dmau_pnrB	8.70	8.43	8.11	8.41	-0.42	1.34
TTEST_pnr	0.123919587					

B'

	Average_Replicates - 120h					
	a	b	c			
Dmel_pnrB	29.27	29.56	30.39			
Dmel_Actin	21.97	21.87	22.00	Average_Δct		
Δct_Dmel_pnrB	7.30	7.68	8.39	7.79		
Dmau_pnrB	28.40	28.50	28.21			
Dmau_Actin	21.17	21.31	21.14	Average_Δct	ΔΔct	2^-ΔΔct
Δct_Dmau_pnrB	7.23	7.19	7.07	7.16	-0.63	1.54
TTEST_pnr	0.18534774					

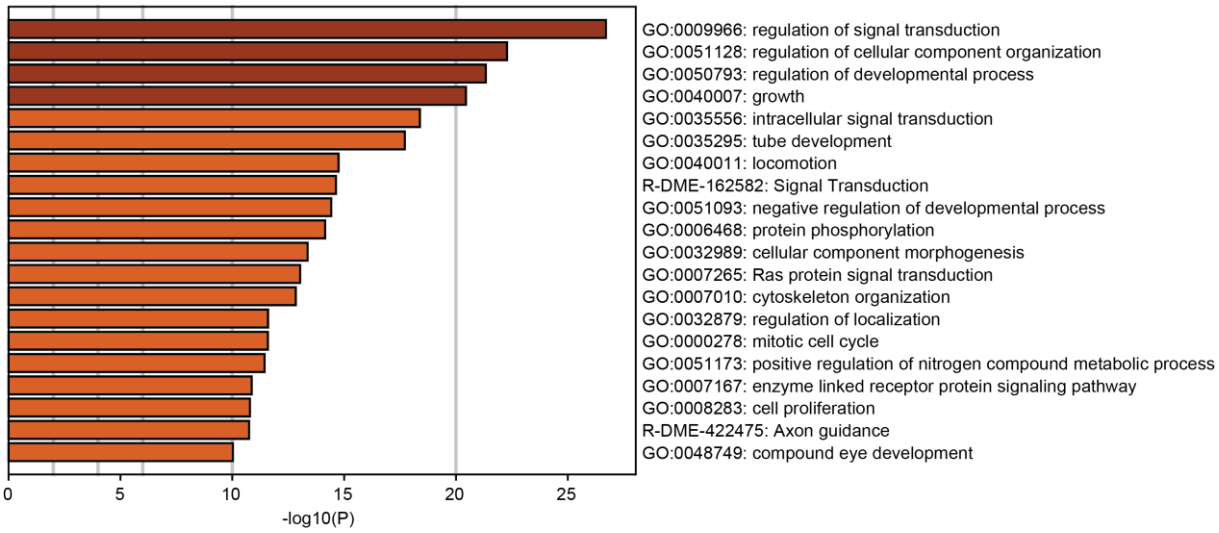
C Average_Δct = ((Dmel_pnrB_a - Dmel_Actin_a) + (Dmel_pnrB_b - Dmel_Actin_b) + (Dmel_pnrB_c - Dmel_Actin_c))/3
ΔΔct = Δct_Dmel_pnrB -Δct_Dmau_pnrB

Supplementary Figure 11. A. q-RT PCR for *pnrB* in *D. melanogaster* and *D. mauritiana*. Efficiency of the used primers in *D. melanogaster* for *pnrB*. we prepared four 1:4 dilutions of a pool of all RNA samples per species (1:2, 1:6, 1:18, 1:54). The primer pair for *pnrB* yielded an efficiency of 98% in *D. melanogaster* and the primer pair for *actin* yielded an efficiency of 106%. **B.** Same efficiency calculations for the same genes as in A for *D. mauritiana*. In this species the primer pair for *pnrB* showed an efficiency of 104% and 100.5% for *actin*. **C.** Comparison of *pnr* expression levels in *D. melanogaster* and *D. mauritiana* at 96h AEL. *actin* was in all cases used as reference gene. Log2fold changes were calculated using the ΔΔCT method. At 96h AEL, expression of *pnr* was 1.3x higher than in *D. melanogaster* at the same time point. **D.** Comparison of *pnr* expression levels in *D. melanogaster* and *D. mauritiana* at 120h AEL. Expression of *pnr* was 1.5x higher than in *D. melanogaster* at the same time point. Even though the difference in expression is not significant we observe the same trend as in the transcriptomics dataset.

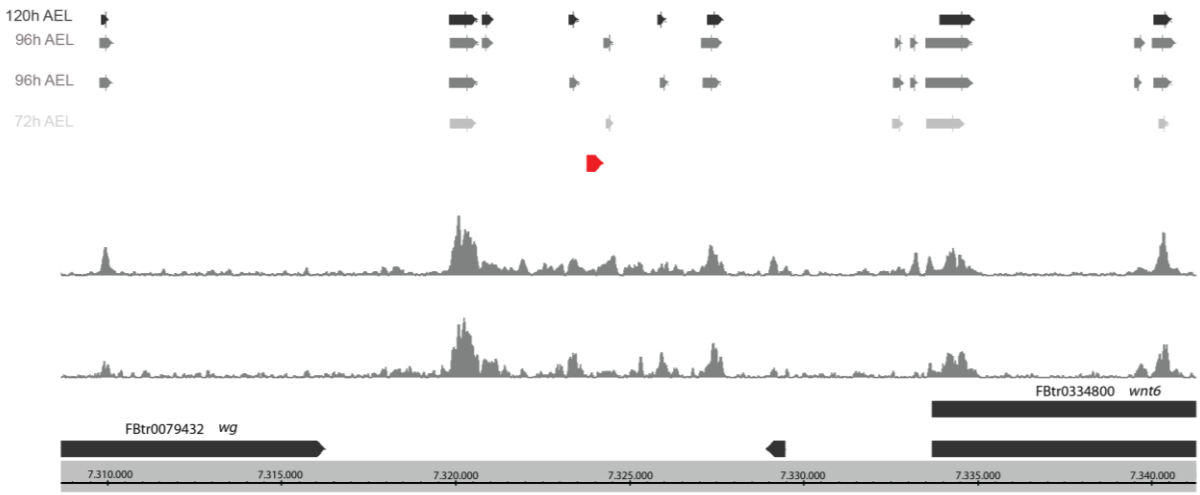
A

possibleTFs	NES
pnr	7.13
nejire	5.71
Pmad	4.75
mef2	3.43

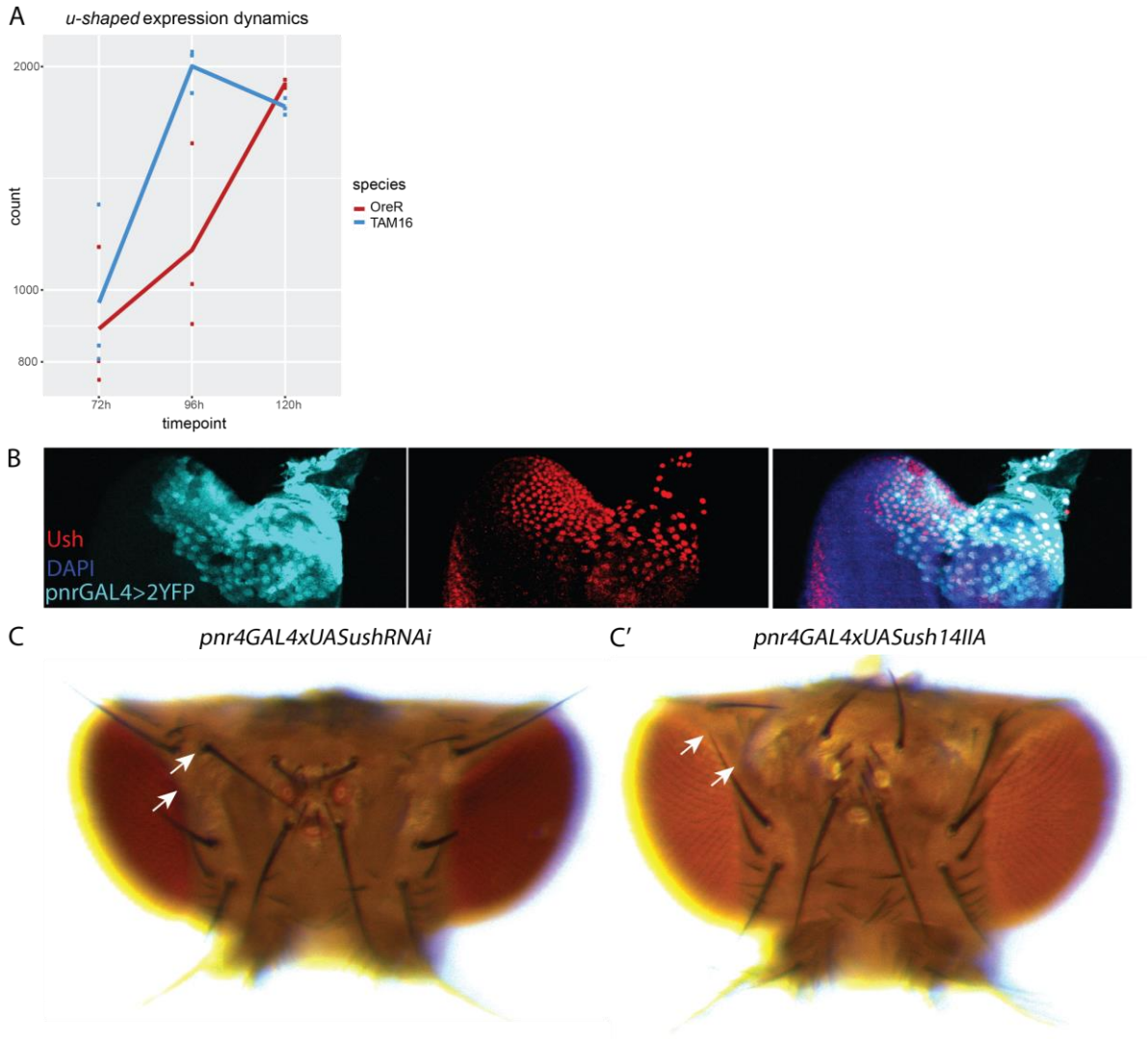
B



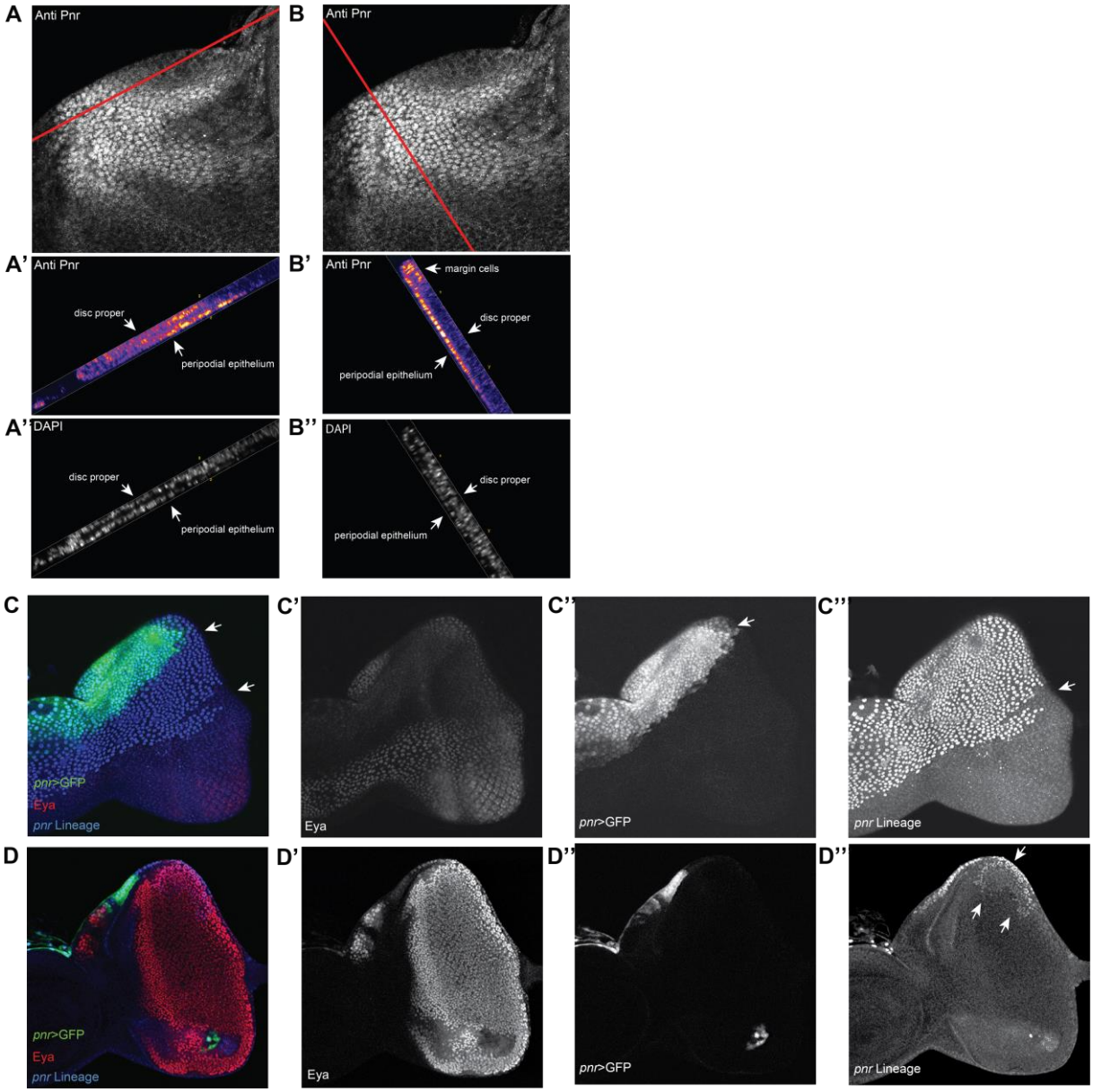
Supplementary Figure 12. A. Cross validation of TFBS enrichment in 5kb upstream of the TSS, 5'UTR regions and 1st introns of all predicted Pnr target genes, NES-values are given in the second column. **B.** GO-term enrichment analysis of all predicted Pnr target genes are enriched in processes like signal transduction, growth, cell cycle but also in more specific terms like compound eye development.



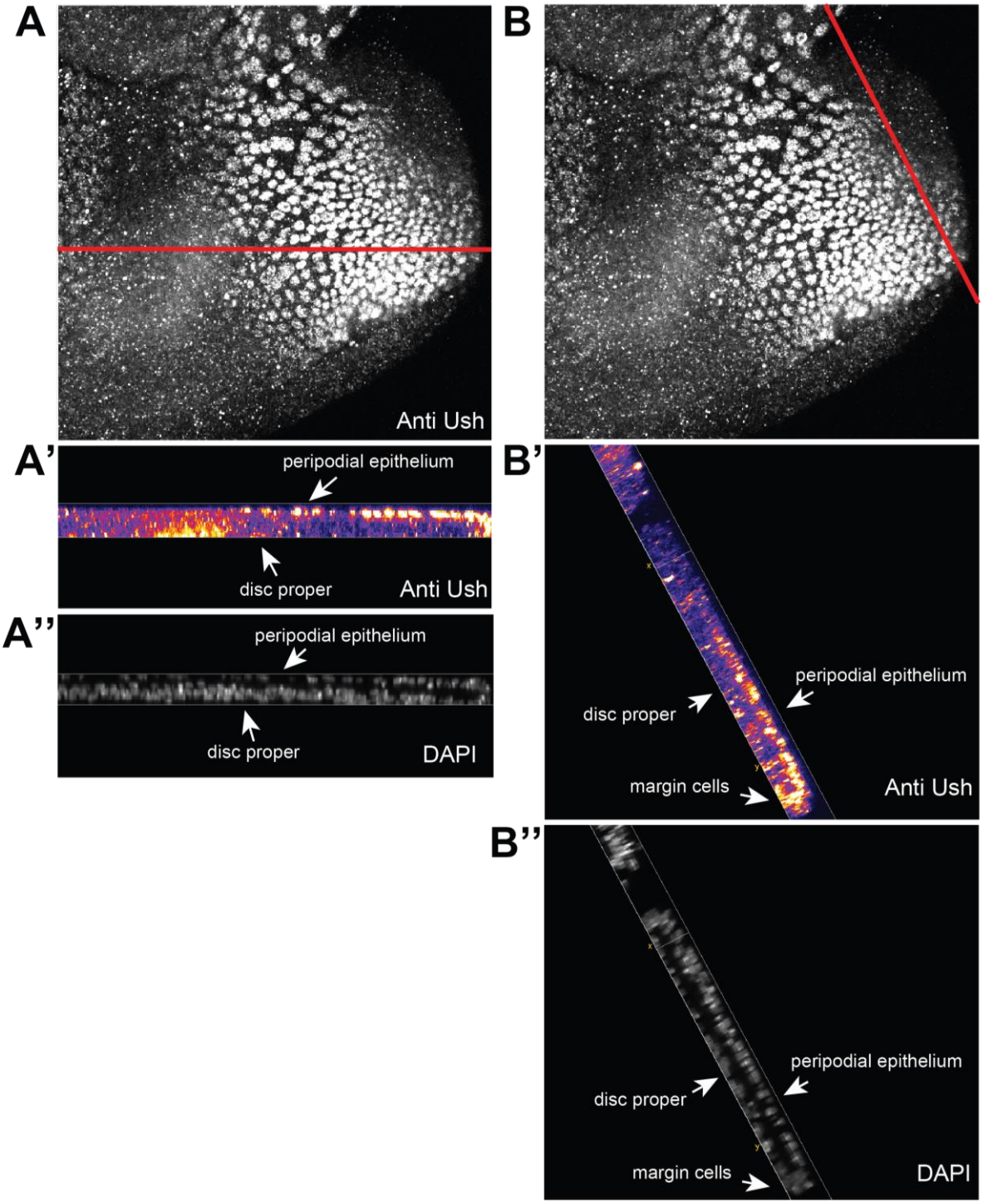
Supplementary Figure 13. Gene locus of *wg* and *wnt6*. The grey tracks show the depth graph of the ATAC-seq dataset at 96h AEL. The grey bars are significantly called peaks at three timepoints (72h AEL – light grey, 96h AEL – grey, 120h AEL – dark grey). The red bar depicts the *wg*-enhancer region containing two conserved GATA motifs, which are though not activated by Pnr to drive expression in the peripodial epithelium.



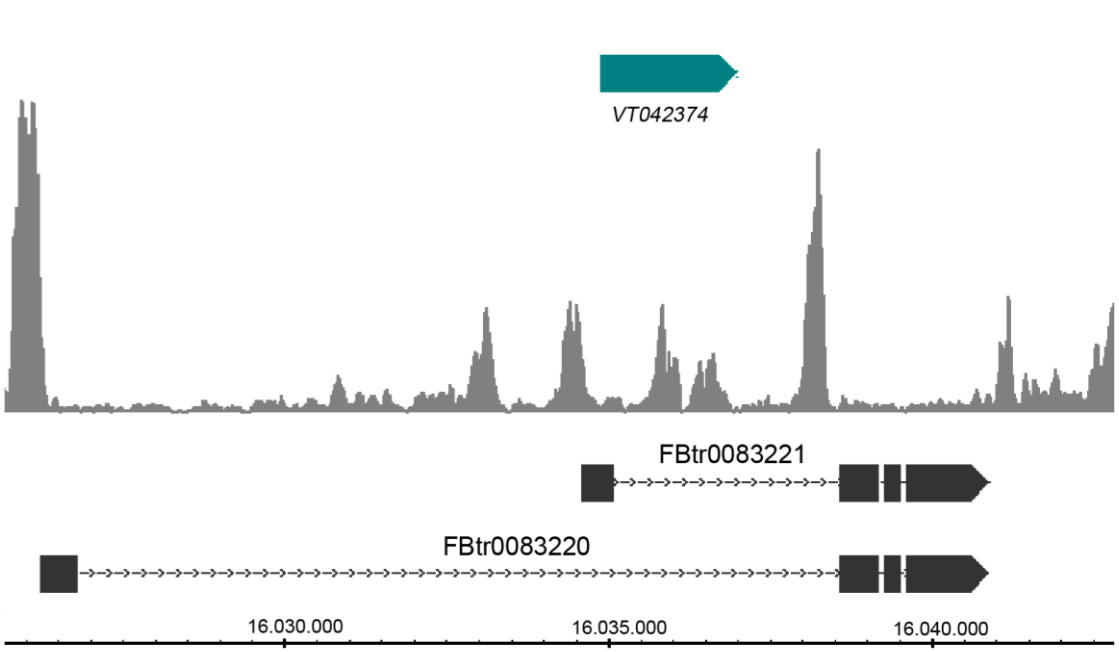
Supplementary Figure 14. A. Expression dynamics of the *ush* transcript in *D. melanogaster* (red) and *D. mauritiana* (blue). **B.** Overlap of *pnr* expression pattern visualized by *pnrGAL4>2YFP* line and Ush protein location, detected with an α-Ush antibody. The two signals overlap in the dorsal part of the developing eye-antennal disc. **C.** Knock-down of *ush* using the *VT042374* driver line and UAS-*ush*RNAi. The head cuticle shows irregularities with loss of the posterior vertical bristles **C'.** Overexpression of *ush* using the *VT042374* driver line and UAS-*ush*14IIA line. As the knock-down of *ush*, upregulation affects the structure of the head cuticle and leads to an overgrowth of the occipital structures. Additionally, the bristle patterns are affected. The posterior vertical bristles as well as the bristles surrounding the eye area are lost.



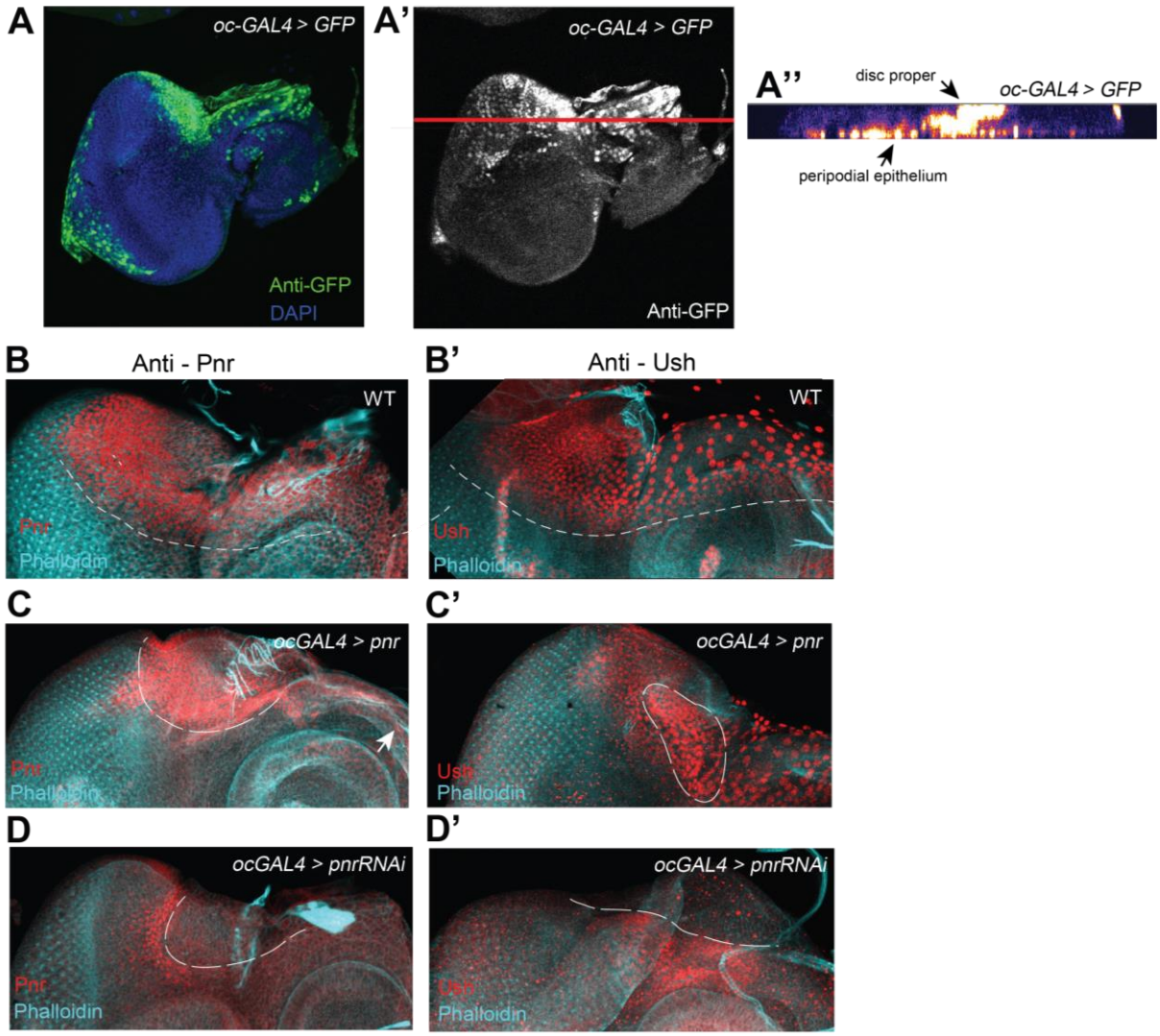
Supplementary Figure 15. A. Pnr protein location (detected with an α-Pnr antibody) in the developing eye-antennal disc at 120h AEL in *D. melanogaster*. **A'.** Vertical section of the same disc as depicted in **A** along the red line showing Pnr antibody staining. Pnr is expressed in the peripodial epithelium and in marginal cells reaching into the disc proper. The intensity of the Pnr signal is lower in the future ocelli region of disc. **A''.** Vertical section of the same disc as in **A** along the red line showing the cell nuclei using DAPI staining. **B.** Pnr protein location (detected with an α-Pnr antibody) in the same disc as in **B**. **B'.** Vertical section of the same disc as depicted in **B** along the red line showing Pnr antibody staining. Pnr is expressed in the peripodial epithelium and in the marginal cells of the disc proper. **B''.** Vertical section of the same disc as in **B** along the red line showing the cell nuclei using DAPI staining. **C-D. Lineage of *pnr*-expressing cells in the developing eye-antennal disc. C-C'''.** *pnr>GFP* expression can be detected in the margin cells (**A''**) of the disc proper. Eyeless expression is shown in red (**A** and **A'**), *pnr*-GFP in green. The *pnr*-lineage shows (in blue) that a view cells that were initially *pnr*-expressing, are forming the dorsal-most cells of the developing retina (**A'''**). **D-D'''.** *pnr>GFP* and the cells of the *pnr*-lineage in the peripodial membrane of the same disc as in **A**. **D.** Overlap of *pnr>GFP* cells (green) and the *pnr*-lineage in blue. Eyeless cannot be detected in the disc proper (**D'**). The *pnr*-driver line drives expression of GFP in the dorsal most region of the developing eye-antennal disc (**D''**). Cells that initially expressed *pnr* cover the complete dorsal lineage of the eye-antennal disc including the retina (**D'''**).



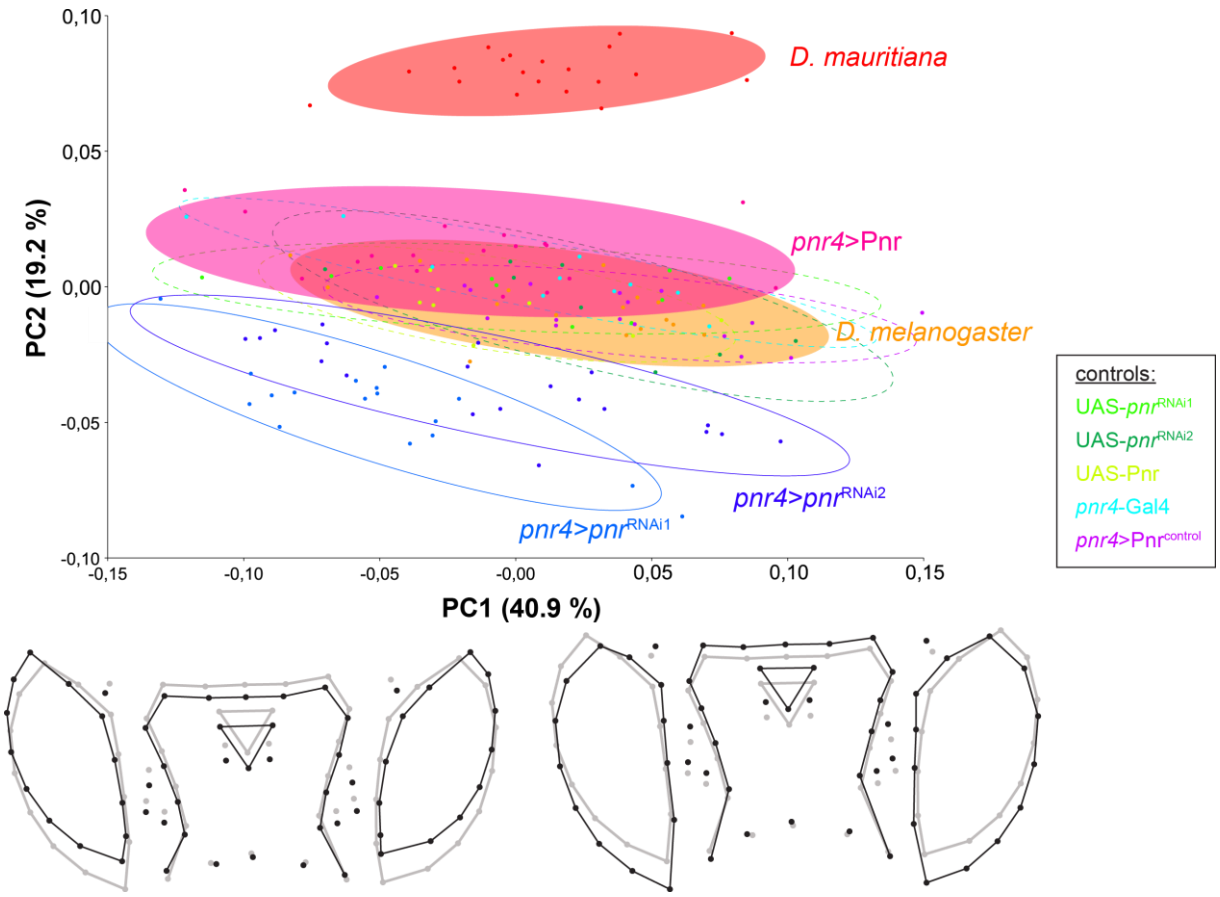
Supplementary Figure 16. A. Ush protein location (detected with an α-Ush antibody) in the developing eye-antennal disc at 120h AEL in *D. melanogaster*. **A'.** Vertical section of the same disc as depicted in **A** along the red line showing Ush antibody staining. Ush is expressed in the peripodial epithelium. **A''.** Vertical section of the same disc as in **B** along the red line showing the cell nuclei using DAPI staining. **B.** Ush protein location (detected with an α-Ush antibody) in the same disc as in **A**. **B'.** Vertical section of the same disc as depicted in **B** along the red line showing Ush antibody staining. Ush is expressed in the peripodial epithelium and in marginal cells reaching into the disc proper. **B''.** Vertical section of the same disc as in **B** along the red line showing the cell nuclei using DAPI staining.



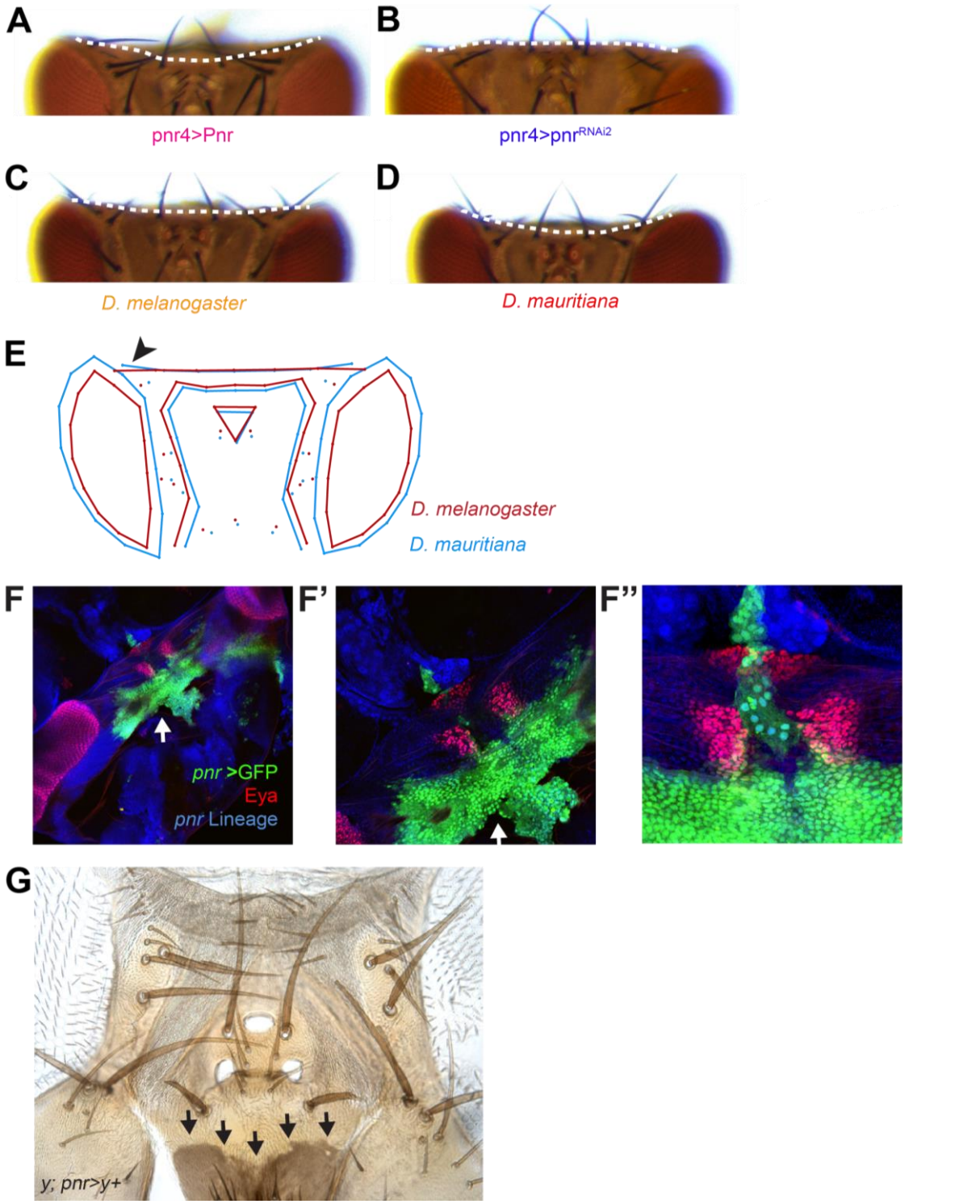
Supplementary Figure 17. *pnr* locus showing two isoforms, namely *pnrA* (FBtr0083221) and *pnrB* (FBtr0083220). The grey track shows the ATAC-seq data at 96h AEL represented as a depth graph. The cyan bar represents the 2kb DNA fragment that controls expression of GAL4 of the *VT042374* driver line. It overlaps with two open-chromatin peaks, which are potential regulatory regions for the expression of *pnr* in the eye-antennal disc.



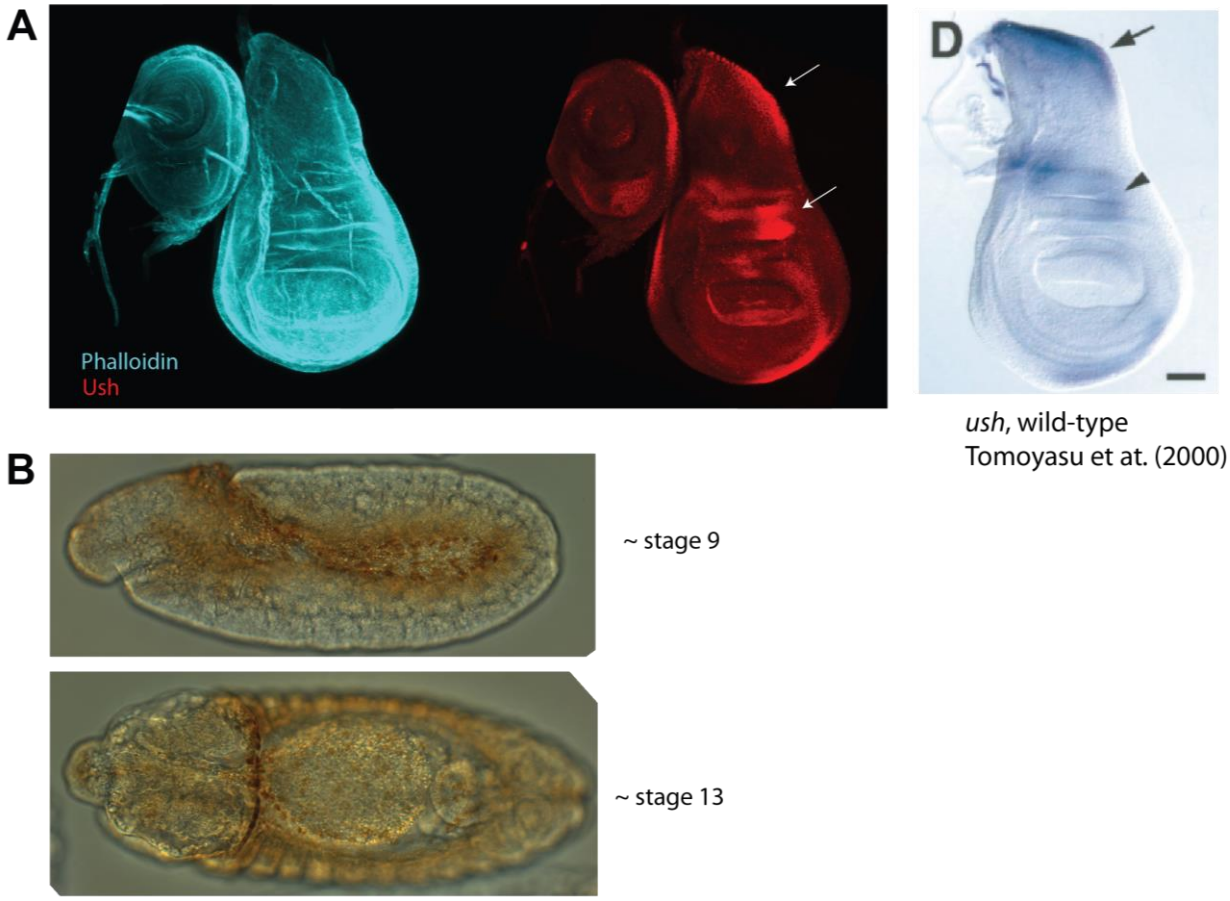
Supplementary Figure 18. A. *oc-GAL4>UAS-GFP* (StingerII lline). **A'.** Vertical section along the red line in the same disc as in **A** showing GFP expression in a few cells of the disc proper and in the peripodial epithelium. **B.** Pnr location in *D. melanogaster* WT eye-antennal disc at 120h AEL, detected with an α-Pnr antibody **B'.** Ush location in *D. melanogaster* WT eye-antennal disc at 120h AEL. **C.** Overexpression of *pnr* using the oc-GAL4 driver line leads to a stronger antibody signal in the future *oc*-region of the developing disc. **C'.** A slight upregulation of Ush signal can be detected in the future *oc*-region of the developing disc, upon upregulation of *pnr* using the *oc*-GAL4 driver line. White dotted lines mark the border between stronger antibody signals were *pnr* is overexpressed and weaker endogenous expression. **D.** Upon knock-down of *pnr* using the *oc*-GAL4 driver line, Pnr antibody staining is lost in the future oc-region. **D'.** Upon knock-down of *pnr* using the *oc*-GAL4 driver line, Ush antibody staining is lost in the future ocelli-region. Note that the dorsal part of the eye-antennal disc is folded in this picture. White dotted lines mark the border where antibody signal still can be detected and where it is lost, due to *pnr* RNAi.



Supplementary Figure 19. Principle component analysis of dorsal head shapes. Shown are PC1 against PC2. Red and orange clouds represent the 'WT-like' head shapes (*D. melanogaster* in orange and *D. mauritiana* in red). Overexpression of *pnr* is represented in pink. The blue empty circles represent knock-down of *pnr*, with a weak effector line in dark blue and a strong effector RNAi line in light blue. The dotted lined circles represent head shapes of the parental UAS- and GAL4 fly lines, that were used to set up the crosses for overexpression and knock-down of *pnr*.



Supplementary Figure 20. A-D. Dorsal-most view of adult heads of *D. melanogaster* and *D. mauritiana* WT flies, *VT042374>pnr* and *VT042374>pnr*^{RNAi2} flies. The white dotted line represents the occipital region, showing the variation in this structure in the different lines: **A**: *VT042374>pnr*; overexpression of *pnr*. **B.** *VT042374>pnr*^{RNAi2} knock-down of *pnr*. **C.** *D. melanogaster* **D.** *D. mauritiana*. **E.** Mean head shapes of *D. melanogaster* and *D. mauritiana* using 64 landmarks (instead of 57) including the occipital region. Discriminant function analysis clearly reveals the convex form of this region in *D. mauritiana* (see black arrowhead). **F-F''.** *pnr*-expression in developing pupal head structures. Cells marked with *pnr>GFP* are accumulating in the future occipital region (green), right behind the developing ocelli (red), and the head region where the two discs are fusing. **D.** The *y*-rescued area representing the *pnr*-domain, moves towards the occipital region (black arrows) in the adult *Drosophila* head.



Supplementary Material Figure 1. A. Ush protein location in the developing *Drosophila* wing disc, detected with the newly generated α-Ush antibody. The regions where Ush can be detected is reminiscent of the region where *ush* mRNA was detected using *in-situ* hybridization in (Tomoyasu et al., 2000) (see white and black arrows). **B.** Ush protein location in the developing *Drosophila* embryo at ~stage 9 and ~stage 13, detected with an α-Ush antibody.

4.7. Supplementary Tables

Supplementary Table 19. List of putative Pnr target genes.

FBgn	GeneSymbol	FBgn	GeneSymbol	FBgn	GeneSymbol	FBgn	GeneSymbol
FBgn0027786	Mtch	FBgn0015795	Rab7	FBgn0030610	CG9065	FBgn0004167	kst
FBgn0016984	sktl	FBgn0264785	Hph	FBgn0012051	CalpA	FBgn0041188	Atx2
FBgn0053111	CG33111	FBgn0052423	shep	FBgn0032901	sky	FBgn0261270	SelD
FBgn0086856	CG11555	FBgn0011586	e(r)	FBgn0029840	raptor	FBgn0263005	CG43313
FBgn0038834	RpS30	FBgn0022029	l(2)k01209	FBgn0038890	CG7956	FBgn0032988	Tif-IA
FBgn0266570	NO66	FBgn0040212	Dhap-at	FBgn0023519	mRpL16	FBgn0039633	CG11873
FBgn0038504	Sur-8	FBgn0011817	nmo	FBgn0004876	cdi	FBgn0044020	Roc2
FBgn0037358	elm	FBgn0014020	Rho1	FBgn0083968	CG34132	FBgn0037363	Atg17
FBgn0046704	Liprin-alpha	FBgn0000611	exd	FBgn0031310	Vps29	FBgn0030341	p24-1
FBgn0036381	CG8745	FBgn0033649	pyr	FBgn0261574	kug	FBgn0015799	Rbf
FBgn0015279	Pi3K92E	FBgn0263144	bin3	FBgn0030396	CG2556	FBgn0020261	pcm
FBgn0003660	Syb	FBgn0038191	CG9925	FBgn0031174	CG1486	FBgn0027597	CG17712
FBgn0029662	CG12206	FBgn0028484	Ack	FBgn0003557	Su(dx)	FBgn0032633	Lrch
FBgn0260748	CG5004	FBgn0000405	CycB	FBgn0015789	Rab10	FBgn0038787	CG4360
FBgn0026418	Hsc70Cb	FBgn0005671	Vha55	FBgn0039929	CG11076	FBgn0051698	CG31698
FBgn0029976	snz	FBgn0261538	CG42662	FBgn0061198	HSPC300	FBgn0035989	CG3967
FBgn0027342	fz4	FBgn0037696	GstZ1	FBgn0000017	Abl	FBgn0003205	Ras85D
FBgn0283477	SF2	FBgn0052479	Usp10	FBgn0033340	CG13751	FBgn0040056	CG17698
FBgn0030502	tth	FBgn0032197	CG5694	FBgn0003371	sgg	FBgn0036970	Spn77Bc
FBgn0025574	Pli	FBgn0034091	mrj	FBgn0005659	Ets98B	FBgn0036257	RhoGAP68F
FBgn0026533	Dek	FBgn0039158	TBC1d7	FBgn0037906	PGRP-LB	FBgn0031850	Tsp
FBgn0015791	Rab14	FBgn0034436	CG11961	FBgn0030030	CG1636	FBgn0031609	CG15443
FBgn0035414	CG14965	FBgn0026206	mei-P26	FBgn0030503	Tango2	FBgn0034503	MED8
FBgn0029944	Dok	FBgn0000711	flw	FBgn0052529	Hers	FBgn0038551	Odj
FBgn0037551	Arl8	FBgn0040087	p115	FBgn0263216	CG43386	FBgn0002715	mei-S332
FBgn0086757	cbs	FBgn0004837	Su(H)	FBgn0037846	CG6574	FBgn0259176	bun
FBgn0263603	Zn72D	FBgn0261524	lic	FBgn0022787	Hel89B	FBgn0000183	BicD
FBgn0050122	CG30122	FBgn0004888	Scsalpha1	FBgn0026196	nop5	FBgn0023143	Uba1
FBgn0035088	CG3776	FBgn0030686	mRpL3	FBgn0002645	Map205	FBgn0028509	CenG1A
FBgn0261609	eIF2alpha	FBgn0029689	CG6428	FBgn0035121	Tudor-SN	FBgn0250843	Prosalpha6
FBgn0053156	CG33156	FBgn0030873	CG15814	FBgn0264712	CG1172	FBgn0262117	IntS3
FBgn0030616	RpL37a	FBgn0036028	CG16717	FBgn0002354	l(3)87Df	FBgn0025936	Eph
FBgn0027866	CG9776	FBgn0039233	CG7006	FBgn0015331	abs	FBgn0001138	grn
FBgn0035540	Syx17	FBgn0267849	Syx7	FBgn0030581	CG14408	FBgn0000286	Cf2
FBgn0011592	fra	FBgn0004370	Ptp10D	FBgn0023212	EloB	FBgn0032817	CG10631
FBgn0025864	Crag	FBgn0263352	Unr	FBgn0000179	bi	FBgn0027532	CG7139
FBgn0037978	KLHL18	FBgn0020279	lig	FBgn0037082	CG5664	FBgn0031988	CG8668
FBgn0011604	Iswi	FBgn0027553	NELF-B	FBgn0038156	side-IV	FBgn0053653	Cadps
FBgn0014879	Set	FBgn0034194	CG15611	FBgn0004177	mts	FBgn0038454	CG10324
FBgn0031253	CG11885	FBgn0033463	CG1513	FBgn0021874	Nle	FBgn0035157	CG13894
FBgn0016977	spen	FBgn0038143	CG9813	FBgn0044826	Pak3	FBgn0028984	Spn88Ea
FBgn0037918	CG6791	FBgn0032725	Nedd8	FBgn0029504	CHES-1-like	FBgn0038662	Mpc1

FBgn0001087	g	FBgn0039140	Miro	FBgn0024734	PRL-1	FBgn0031769	CG9135
FBgn0041087	wun2	FBgn0032482	Pect	FBgn0000015	Abd-B	FBgn0040344	CG3711
FBgn0039641	CG14511	FBgn0024314	Plap	FBgn0029801	CG15771	FBgn0017581	Lk6
FBgn0031263	Tspo	FBgn0039266	CG11791	FBgn0004395	unk	FBgn0002774	mle
FBgn0003317	sax	FBgn0052177	Ndfip	FBgn0036926	CG7646	FBgn0030435	CG4645
FBgn0061476	Zwilch	FBgn0002973	numb	FBgn0052708	CG32708	FBgn0001218	Hsc70-3
FBgn0003447	sn	FBgn0031474	CG2991	FBgn0038947	Sar1	FBgn0041706	CG3253
FBgn0259202	CG42306	FBgn0032614	CG13284	FBgn0259113	DNApol-alpha180	FBgn0004587	B52
FBgn0039068	CG13827	FBgn0002643	mam	FBgn0027872	rdgBbeta	FBgn0001995	mRpL4
FBgn0002775	msl-3	FBgn0039654	Brd8	FBgn0035159	CG13896	FBgn0016131	Cdk4
FBgn0020653	Trxr-1	FBgn0033229	CG12822	FBgn0261550	CG42668	FBgn0003423	slgA
FBgn0037710	CG9393	FBgn0038928	Fadd	FBgn0026375	RhoGAPp190	FBgn0030963	CG7101
FBgn0002431	hyd	FBgn0031023	CG14200	FBgn0030554	CG1434	FBgn0264089	sli
FBgn0015037	Cyp4p1	FBgn0031143	CG1532	FBgn0250838	roh	FBgn0039773	CG2224
FBgn0261986	RASSF8	FBgn0005558	ey	FBgn0051151	wge	FBgn0030243	CG2186
FBgn0015477	Rme-8	FBgn0085370	Pde11	FBgn0013272	Gp150	FBgn0003231	ref(2)P
FBgn0038981	CG5346	FBgn0038053	CG18549	FBgn0028688	Rpn7	FBgn0031053	CG14223
FBgn0038870	Oga	FBgn0039764	CG15535	FBgn0033762	ZnT49B	FBgn0031768	CG12393
FBgn0010303	hep	FBgn0040283	SMC1	FBgn0015774	NetB	FBgn0040237	bor
FBgn0039508	CG3368	FBgn0040660	CG13551	FBgn0027055	CSN3	FBgn0027779	VhaSFD
FBgn0024889	Kap-alpha1	FBgn0031992	Acbp1	FBgn0016693	Past1	FBgn0267252	Ggamma30A
FBgn0050338	CG30338	FBgn0262733	Src64B	FBgn0037188	CG7369	FBgn0039737	CG7920
FBgn0036890	CG9368	FBgn0051126	CG31126	FBgn0037021	CG11399	FBgn0022764	Sin3A
FBgn0021895	ytr	FBgn0051915	CG31915	FBgn0000536	eas	FBgn0035094	CG9380
FBgn0031118	RhoGAP19D	FBgn0035449	CG14971	FBgn0041111	lilli	FBgn0262517	Ltn1
FBgn0015622	Cnx99A	FBgn0003002	opa	FBgn0261244	inaE	FBgn0037841	CG4565
FBgn0039904	Hcf	FBgn0037468	CG1943	FBgn0086359	Invadolysin	FBgn0039226	Ude
FBgn0015229	glec	FBgn0029887	CG3198	FBgn0033052	SCAP	FBgn0038578	MED17
FBgn0003525	stg	FBgn0000319	Chc	FBgn0037900	CG5276	FBgn0032223	GATAd
FBgn0038745	CG4538	FBgn0036913	Usp32	FBgn0039938	Sox102F	FBgn0263258	chas
FBgn0026257	cav	FBgn0033961	ND-B15	FBgn0001105	Gbeta13F	FBgn0260962	pic
FBgn0013954	Fkbp12	FBgn0003301	rut	FBgn0003415	skd	FBgn0038256	CG7530
FBgn0030420	CG12717	FBgn0033089	CG17266	FBgn0033921	tej	FBgn0032919	CG9253
FBgn0086784	stmA	FBgn0051075	CG31075	FBgn0283473	S6KL	FBgn0038826	Syp
FBgn0013305	Nmda1	FBgn0263993	CG43736	FBgn0030082	HP1b	FBgn0004657	mys
FBgn0026373	RpII33	FBgn0037669	Ibf2	FBgn0034674	CG9304	FBgn0028476	Usp1
FBgn0260632	dl	FBgn0262656	Myc	FBgn0034853	Ice1	FBgn0030572	mRpS25
FBgn0034573	CG3295	FBgn0034878	pita	FBgn0030786	mRpL22	FBgn0051436	CG31436
FBgn0262614	pyd	FBgn0037012	Rcd2	FBgn0023529	CG2918	FBgn0038953	CG18596
FBgn0039160	CG5510	FBgn0032943	Tsp39D	FBgn0030478	CG1640	FBgn0266696	Svil
FBgn0085377	CG34348	FBgn0041342	Pcyt1	FBgn0030744	CG9992	FBgn0265630	sno
FBgn0052133	Ptip	FBgn0028375	heix	FBgn0283499	InR	FBgn0020655	ArfGAP1
FBgn0027280	l(1)G0193	FBgn0259749	mmy	FBgn0038686	CG5555	FBgn0265434	zip
FBgn0001124	Got1	FBgn0087013	Karybeta3	FBgn0261799	dsx-c73A	FBgn0027621	Pfrx
FBgn0283724	Girdin	FBgn0031779	CG9175	FBgn0004401	Pep	FBgn0035824	CG8281
FBgn0028662	VhaPPA1-1	FBgn0024973	CG2701	FBgn0026179	siz	FBgn0033127	Tsp42Ef

FBgn0034997	CG3376	FBgn0029820	CG16721	FBgn0036697	rogdi	FBgn0010621	CCT5
FBgn0033770	wuc	FBgn0037614	TMEM216	FBgn0004903	Rb97D	FBgn0261439	SdhA
FBgn0264307	orb2	FBgn0030403	CG1824	FBgn0003274	RpLP2	FBgn0260970	Ubr3
FBgn0031449	CG31689	FBgn0000810	fs(1)K10	FBgn0003396	shn	FBgn0025839	ND-B14.5A
FBgn0038659	EndoA	FBgn0030505	NFAT	FBgn0034657	LBR	FBgn0035357	MEP-1
FBgn0039155	Kal1	FBgn0035907	GstO1	FBgn0030973	CG7332	FBgn0039966	Rab21
FBgn0028897	CG4935	FBgn0039215	CG6695	FBgn0038877	CG3308	FBgn0031505	ND-B14.5B
FBgn0031574	TTLL4B	FBgn0261647	Axud1	FBgn0005648	Pabp2	FBgn0284250	Oaz
FBgn0028541	TM9SF4	FBgn0028695	Rpn1	FBgn0003134	Pp1alpha-96A	FBgn0038471	CG5220
FBgn0013983	imd	FBgn0037944	CG6923	FBgn0025394	inc	FBgn0003165	pum
FBgn0052141	saturn	FBgn0037874	Tctp	FBgn0033951	CG10139	FBgn0011606	Klp3A
FBgn0031549	Spindly	FBgn0010265	RpS13	FBgn0030956	CG18259	FBgn0028968	gammaCOP
FBgn0004913	Gnf1	FBgn0020309	crol	FBgn0015623	Cpr	FBgn0267791	HnRNP-K
FBgn0031183	CG14621	FBgn0016685	Nlp	FBgn0034914	CG5554	FBgn0027492	wdb
FBgn0002590	RpS5a	FBgn0028408	Drep2	FBgn0031682	CG5828	FBgn0261823	Asx
FBgn0266084	Fhos	FBgn0024754	Flo1	FBgn0259876	Cap-G	FBgn0031681	Pgant5
FBgn0031161	CG15445	FBgn0265192	Snp	FBgn0030055	CG12772	FBgn0029903	pod1
FBgn0032821	CdGAPr	FBgn0002873	mud	FBgn0043903	dome	FBgn0022153	l(2)k05819
FBgn0036373	Tgi	FBgn0051683	CG31683	FBgn0030930	Pgant7	FBgn0029708	CG3556
FBgn0029709	CHOp24	FBgn0061200	Nup153	FBgn0086361	alph	FBgn0025741	PlexA
FBgn0010348	Arf79F	FBgn0039213	atl	FBgn0010488	NAT1	FBgn0000479	dnc
FBgn0051992	gw	FBgn0035909	ergic53	FBgn0002638	Rcc1	FBgn0017567	ND-23
FBgn0260780	wisp	FBgn0025185	az2	FBgn0265784	CrebB	FBgn0027343	fz3
FBgn0001169	H	FBgn0031250	Ent1	FBgn0265052	St3	FBgn0011272	RpL13
FBgn0032656	CG5674	FBgn0010408	RpS9	FBgn0033673	CG8298	FBgn0000404	CycA
FBgn0037912	sea	FBgn0031144	CG1529	FBgn0003159	CG2841	FBgn0001941	ifc
FBgn0015024	CkIalpha	FBgn0039705	Atg16	FBgn0039590	CG10011	FBgn0025724	beta'COP
FBgn0013749	Arf102F	FBgn0039851	mey	FBgn0028360	Cdc7	FBgn0036974	eRF1
FBgn0028717	Lnk	FBgn0036762	CG7430	FBgn0034528	CG11180	FBgn0035689	CG7376
FBgn0265140	Meltrin	FBgn0033154	CG1850	FBgn0033199	CG17985	FBgn0265298	SC35
FBgn0263987	spoon	FBgn0010198	RpS15Aa	FBgn0010113	hdc	FBgn0034240	MESR4
FBgn0022213	Cse1	FBgn0000473	Cyp6a2	FBgn0037238	CG1090	FBgn0024555	flfl
FBgn0030693	CG8974	FBgn0015320	Ubc2	FBgn0039635	Pdhb	FBgn0037073	Tsr1
FBgn0039026	CG7029	FBgn0022960	vimar	FBgn0051989	Cap-D3	FBgn0037234	CG9795
FBgn0039338	XNP	FBgn0034570	CG10543	FBgn0025382	Rab27	FBgn0028331	l(1)G0289
FBgn0023130	a6	FBgn0004907	14-3-3zeta	FBgn0053469	CG33469	FBgn0030592	CG9514
FBgn0037354	CG12171	FBgn0052280	CG32280	FBgn0263392	Tet	FBgn0027339	jim
FBgn0035558	CG11357	FBgn0264975	Nrg	FBgn0039136	CG5902	FBgn0025335	Cpes
FBgn0025681	CG3558	FBgn0015270	Orc2	FBgn0028425	JhI-21	FBgn0032029	CG17292
FBgn0001075	ft	FBgn0041585	olf186-F	FBgn0011656	Mef2	FBgn0283657	Tlk
FBgn0263231	bel	FBgn0034742	CG4294	FBgn0029679	CG2901	FBgn0086694	Bre1
FBgn0034230	CG4853	FBgn0001269	inv	FBgn0005655	PCNA	FBgn0031037	CG14207
FBgn0024184	unc-4	FBgn0038197	foxo	FBgn0003139	PpV	FBgn0036828	CG6841
FBgn0024733	RpL10	FBgn0031145	Ntf-2	FBgn0039665	CG2310	FBgn0028693	Rpn12
FBgn0002044	swm	FBgn0015218	eIF4E1	FBgn0039908	Asator	FBgn0032006	Pvr
FBgn0038320	Sra-1	FBgn0035046	ND-19	FBgn0032363	Dlg5	FBgn0041186	Slbp
FBgn0261885	osa	FBgn0267912	CanA-14F	FBgn0264294	Cyt-b5	FBgn0032475	Sfmbt

FBgn0031717	Oscillin	FBgn0250786	Chd1	FBgn0003391	shg	FBgn0032339	Wdr59
FBgn0034313	CG5726	FBgn0025830	IntS8	FBgn0000719	fog	FBgn0036566	ClC-c
FBgn0020412	JIL-1	FBgn0283536	Vha13	FBgn0003042	Pc	FBgn0024909	Taf7
FBgn0021760	chb	FBgn0003701	thr	FBgn0003189	r	FBgn0050372	Asap
FBgn0085693	CG41562	FBgn0029006	Smurf	FBgn0052212	CG32212	FBgn0039920	CG11360
FBgn0001139	gro	FBgn0010583	dock	FBgn0031883	Caper	FBgn0024251	bbx
FBgn0036449	bmm	FBgn0035452	CG10359	FBgn0029878	Pat1	FBgn0038755	Hs6st
FBgn0030349	CG10353	FBgn0260858	Ykt6	FBgn0030049	Trf4-1	FBgn0029095	aru
FBgn0038350	AOX4	FBgn0262114	RanBPM	FBgn0262734	eIF4H1	FBgn0032075	Tsp29Fb
FBgn0029990	CG2233	FBgn0037282	CG14657	FBgn0028394	CG17834	FBgn0033925	CG8617
FBgn0001301	kel	FBgn0040208	Kat60	FBgn0261268	Cul3	FBgn0032848	nesd
FBgn0004655	wapl	FBgn0036476	sstn	FBgn0051108	TTLL5	FBgn0003716	tkv
FBgn0030366	Usp7	FBgn0029825	CG12728	FBgn0033715	CG8490	FBgn0036240	CG6928
FBgn0011754	PhKgamma	FBgn0003731	Egfr	FBgn0042134	Capr	FBgn0036964	FRG1
FBgn0037092	M6	FBgn0053181	CG33181	FBgn0031314	IntS14	FBgn0010620	CG10939
FBgn0037549	CG7878	FBgn0260462	CG12163	FBgn0034418	CG15118	FBgn0085430	CG34401
FBgn0283741	prage	FBgn0000581	E(Pc)	FBgn0034488	Hacl	FBgn0052756	CG32756
FBgn0037561	CG9630	FBgn0029685	CG2938	FBgn0261444	CG3638	FBgn0000261	Cat
FBgn0038853	RhoGAP93B	FBgn0015790	Rab11	FBgn0260400	elav	FBgn0031881	MME1
FBgn0004959	phm	FBgn0083984	CG34148	FBgn0041174	Vhl	FBgn0001189	hfw
FBgn0029629	eIF3g1	FBgn0000394	cv	FBgn0037814	CG6325	FBgn0035533	Cip4
FBgn0020440	Fak	FBgn0038976	Pfdn5	FBgn0037391	CG2017	FBgn0035120	wac
FBgn0030996	CG14194	FBgn0039852	nyo	FBgn0034854	Golgin245	FBgn0024957	Irp-1B
FBgn0023458	Rbcn-3A	FBgn0036846	MESR6	FBgn0039065	Rad60	FBgn0033519	CG11825
FBgn0034763	RYBP	FBgn0265082	Cdep	FBgn0031036	CG14220	FBgn0019637	Atu
FBgn0028506	CG4455	FBgn0004864	hop	FBgn0261019	moi	FBgn0029958	Pdp
FBgn0031492	CG3542	FBgn0267326	Ntl	FBgn0027932	Akap200	FBgn0034225	veil
FBgn0035148	CG3402	FBgn0052831	CG33695	FBgn0028494	CG6424	FBgn0004397	Vinc
FBgn0028474	CG4119	FBgn0261705	CG42741	FBgn0035001	Slik	FBgn0019968	Khc-73
FBgn0030912	CG6023	FBgn0083167	Neb-cGP	FBgn0053293	CG33293	FBgn0037138	P5CDh1
FBgn0003117	pnr	FBgn0015524	otp	FBgn0260789	mxc	FBgn0266111	ana3
FBgn0033774	CG12374	FBgn0010328	woc	FBgn0014388	sty	FBgn0267390	dop
FBgn0020249	stck	FBgn0262127	kibra	FBgn0263705	Myo10A	FBgn0052758	CG32758
FBgn0259168	mnb	FBgn0025615	Torsin	FBgn0032444	CCT4	FBgn0035890	CG13667
FBgn0265003	koi	FBgn0025628	CG4199	FBgn0021856	l(2)k14505	FBgn0037218	aux
FBgn0262743	Fs(2)Ket	FBgn0026259	eIF5B	FBgn0052767	CG32767	FBgn0034708	Vps35
FBgn0001978	stc	FBgn0031126	Cyp6v1	FBgn0030869	Socs16D	FBgn0085220	Ufm1
FBgn0052473	CG32473	FBgn0031488	CG17265	FBgn0261722	fwe	FBgn0025637	SkpA
FBgn0025455	CycT	FBgn0032341	Reps	FBgn0086906	sls	FBgn0037525	CG17816
FBgn0067622	LSm-4	FBgn0031044	MKP-4	FBgn0035713	velo	FBgn0262740	Evi5
FBgn0004103	Pp1-87B	FBgn0000283	Cp190	FBgn0032899	CG9338	FBgn0051184	LSm3
FBgn0038737	CG11447	FBgn0034061	Ufc1	FBgn0017418	ari-1	FBgn0026376	Rgl
FBgn0003310	S	FBgn0039914	mav	FBgn0025693	ZnT41F	FBgn0021825	DCTN2-p50
FBgn0086368	tw	FBgn0035644	DNApol-epsilon58	FBgn0040309	Jafrac1	FBgn0015772	Nak
FBgn0039265	CG11790	FBgn0039733	CG11504	FBgn0037892	mRpL40	FBgn0015379	dod
FBgn0041604	dlp	FBgn0261570	CG42684	FBgn0015754	Lis-1	FBgn0263108	BtbVII

FBgn0025885	Inos	FBgn0033033	scaf	FBgn0259481	Mob2	FBgn0019925	Surf4
FBgn0033259	CG11210	FBgn0260442	rhea	FBgn0033766	Nup188	FBgn0263257	Cngl
FBgn0038769	CG10889	FBgn0029929	CG4593	FBgn0035432	ZnT63C	FBgn0001248	Idh
FBgn0035213	CG2199	FBgn0025633	CG13366	FBgn0000259	CkIIbeta	FBgn0039664	CG2006
FBgn0035617	l(3)psg2	FBgn0030327	FucT6	FBgn0043900	pygo	FBgn0087035	AGO2
FBgn0036008	CG3408	FBgn0030121	Cfp1	FBgn0010531	Ccs	FBgn0033005	CG3107
FBgn0035771	Sec63	FBgn0027951	MTA1-like	FBgn0010808	Chchd3	FBgn0033480	mRpL42
FBgn0013987	MAPk-Ak2	FBgn0020312	Tmtc3	FBgn0002542	lds	FBgn0031540	Pif1
FBgn0032833	COX4	FBgn0024556	mEFTu1	FBgn0265193	Atf-2	FBgn0036340	SRm160
FBgn0053265	Muc68E	FBgn0015621	Clp	FBgn0011224	heph	FBgn0035388	CG2162
FBgn0003891	tud	FBgn0265101	Sgt1	FBgn0052816	CG32816	FBgn0016926	Pino
FBgn0030218	CG1628	FBgn0022699	D19B	FBgn0260392	CG42518	FBgn0052772	CG32772
FBgn0035574	RhoGEF64C	FBgn0066084	RpL41	FBgn0015269	Nf1	FBgn0035630	CG10576
FBgn0028687	Rpt1	FBgn0052425	CG32425	FBgn0263395	hppy	FBgn0064766	CG7600
FBgn0033844	bbc	FBgn0260450	CalpC	FBgn0001491	l(1)10Bb	FBgn0039680	Cap-D2
FBgn0016917	Stat92E	FBgn0262126	Sec24CD	FBgn0003392	shi	FBgn0035356	CG16986
FBgn0261573	CoRest	FBgn0036309	Hip1	FBgn0029763	Usp16-45	FBgn0027066	Eb1
FBgn0002283	l(3)73Ah	FBgn0263237	Cdk7	FBgn0034049	bdg	FBgn0034527	CG9945
FBgn0034925	CG5339	FBgn0085451	htk	FBgn0027865	Tsp96F	FBgn0000289	cg
FBgn0032751	CG17343	FBgn0086372	lap	FBgn0011336	Stt3B	FBgn0034500	CG11200
FBgn0041789	Pax	FBgn0045038	Proc	FBgn0001316	klar	FBgn0266717	Bruce
FBgn0031150	bves	FBgn0052296	Mrtf	FBgn0034542	Fem-1	FBgn0031874	CG13775
FBgn0024998	CG2685	FBgn0040251	Ugt302K1	FBgn0261934	dikar	FBgn0062413	Ctr1A
FBgn0037998	Cog1	FBgn0043884	mask	FBgn0050389	CG30389	FBgn0030141	Gga
FBgn0026379	Pten	FBgn0261477	slim	FBgn0038269	Rrp6	FBgn0030809	Ubr1
FBgn0035640	mad2	FBgn0038145	Droj2	FBgn0262527	nsl1	FBgn0003575	su(sable)
FBgn0031635	tank	FBgn0039741	CG7943	FBgn0283500	Sac1	FBgn0035480	CG14984
FBgn0000617	e(y)1	FBgn0034504	CG8929	FBgn0003079	Raf	FBgn0032050	CG13096
FBgn0263968	nonC	FBgn0051064	CG31064	FBgn0261456	hpo	FBgn0266525	CG45092
FBgn0040493	grsm	FBgn0005585	Calr	FBgn0260743	GC1	FBgn0024846	p38b
FBgn0036575	CG5157	FBgn0015371	chn	FBgn0005427	ewg	FBgn0031575	Cep97
FBgn0261599	RpS29	FBgn0026679	IntS4	FBgn0011760	ctp	FBgn0035024	CG11414
FBgn0035942	ValRS-m	FBgn0000635	Fas2	FBgn0036052	CG10809	FBgn0266917	Sf3a1
FBgn0016983	smid	FBgn0037215	beta-Man	FBgn0052672	Atg8a	FBgn0026149	BCL7-like
FBgn0053062	CG33062	FBgn0027505	Rab3-GAP	FBgn0052264	CG32264	FBgn0243486	rdo
FBgn0035719	tow	FBgn0024432	Dlc90F	FBgn0035087	CG2765	FBgn0015737	Hmu
FBgn0037720	CG8312	FBgn0003204	ras	FBgn0052451	SPoCk	FBgn0086676	spin
FBgn0036376	Liprin-beta	FBgn0039857	RpL6	FBgn0028686	Rpt3	FBgn0039293	Alg9
FBgn0051324	CG31324	FBgn0039209	REPTOR	FBgn0003360	sesB	FBgn0011573	Cdc37
FBgn0022382	Pka-R2	FBgn0037270	eIF3f1	FBgn0026316	Ubc10	FBgn0051457	CG31457
FBgn0004107	Cdk2	FBgn0040334	Tsp3A	FBgn0261108	Atg13	FBgn0028916	CG33090
FBgn0010497	dmGlut	FBgn0003261	Rm62	FBgn0031090	Rab35	FBgn0032378	CycY
FBgn0030674	CG8184	FBgn0001197	His2Av	FBgn0030360	CG1806	FBgn0028689	Rpn6
FBgn0004435	Galphaq	FBgn0024811	Crk	FBgn0083969	CG34133	FBgn0037702	CG8176
FBgn0010288	Uch	FBgn0267698	Pak	FBgn0040239	bc10	FBgn0034067	CG8399
FBgn0000625	eyg	FBgn0051217	modSP	FBgn0012037	Ance	FBgn0026428	HDAC6
FBgn0039620	wat	FBgn0036932	CG14184	FBgn0033929	Tfb1	FBgn0039835	mRpL32

FBgn0045866	bai	FBgn0030808	RhoGAP15B	FBgn0036483	CG12316	FBgn0086855	CG17078
FBgn0037135	CG7414	FBgn0002936	ninaA	FBgn0259483	Mob4	FBgn0004652	fru
FBgn0036341	Syx13	FBgn0028582	lqf	FBgn0264075	tgo	FBgn0020626	Osbp
FBgn0030519	CG11151	FBgn0004870	bab1	FBgn0033261	udd	FBgn0058191	CG40191
FBgn0031736	CG11030	FBgn0024236	foi	FBgn0038055	trus	FBgn0011771	Hem
FBgn0037911	CG10898	FBgn0031420	Atxn7	FBgn0019662	qm	FBgn0039932	fuss
FBgn0033757	muskelin	FBgn0034728	rad50	FBgn0028504	CG12182	FBgn0000044	Act57B
FBgn0035497	CG14995	FBgn0261954	east	FBgn0025638	Roc1a	FBgn0032886	CG9328
FBgn0037440	CRAT	FBgn0022349	CG1910	FBgn0261985	Ptpmeg	FBgn0051279	CG31279
FBgn0031320	CG5126	FBgn0026252	msk	FBgn0029935	CG4615	FBgn0036058	CG6707
FBgn0030753	rngo	FBgn0039623	CG1951	FBgn0033668	exp	FBgn0031374	Wdr62
FBgn0024807	DIP1	FBgn0052262	CG32262	FBgn0040230	dbo	FBgn0283468	slmb
FBgn0263974	qin	FBgn0000721	for	FBgn0260653	serp	FBgn0034735	CG4610
FBgn0032330	Samuel	FBgn0086704	stops	FBgn0267347	squ	FBgn0004875	enc
FBgn0030114	CG17754	FBgn0037220	CG14641	FBgn0250830	CG12547	FBgn0038872	Nelf-A
FBgn0267975	vib	FBgn0028956	mthl3	FBgn0029849	Efr	FBgn0037831	Cap-H2
FBgn0037743	CG8412	FBgn0002921	Atpalpha	FBgn0004050	z	FBgn0035147	Gale
FBgn0263106	DnaJ-1	FBgn0037680	pasi2	FBgn0011236	ken	FBgn0030065	CG12075
FBgn0001994	crp	FBgn0031244	CG11601	FBgn0260659	CG42542	FBgn0035443	CG12010
FBgn0012058	Cdc27	FBgn0051122	CG31122	FBgn0040238	Best1	FBgn0034722	Rtf1
FBgn0037679	Aduk	FBgn0035101	p130CAS	FBgn0039487	gb	FBgn0015396	jumu
FBgn0025743	mbt	FBgn0030648	CG6340	FBgn0000163	baz	FBgn0030137	CG15317
FBgn0034180	Ehbp1	FBgn0026160	tna	FBgn0032774	CG17549	FBgn0030321	CG1703
FBgn0027561	CG18659	FBgn0034475	Obp56h	FBgn0261592	RpS6	FBgn0039668	Trc8
FBgn0026879	CG13364	FBgn0265998	Doa	FBgn0031873	Gas41	FBgn0035393	CG16753
FBgn0259704	Nsun5	FBgn0016754	sba	FBgn0036254	CG5645	FBgn0030089	AP-1gamma
FBgn0032715	CG17597	FBgn0038956	CAH8	FBgn0266918	CG32486	FBgn0030354	Upf1
FBgn0025639	Hmt4-20	FBgn0025463	Bap60	FBgn0051739	AspRS-m	FBgn0035016	CG4612
FBgn0032957	CG2225	FBgn0027497	Madm	FBgn0034530	Rcd6	FBgn0036039	Naa60
FBgn0050421	Usp15-31	FBgn0016694	Pdp1	FBgn0032815	CG10462	FBgn0033549	mms4
FBgn0026250	eIF1A	FBgn0032949	Lamp1	FBgn0026239	gukh	FBgn0026086	Adar
FBgn0028684	Rpt5	FBgn0038039	CG5196	FBgn0004654	Pgd	FBgn0034918	Pym
FBgn0035021	CG4622	FBgn0015278	Pi3K68D	FBgn0283509	Phm	FBgn0037856	Leash
FBgn0046687	Tre1	FBgn0033100	CG3420	FBgn0260484	HIP	FBgn0035270	CG13933
FBgn0040964	CG18661	FBgn0005631	robo1	FBgn0053113	Rtnl1	FBgn0086613	Ino80
FBgn0085481	CG34452	FBgn0036684	CG3764	FBgn0029893	CG14442	FBgn0028325	Pdha
FBgn0027546	CG4766	FBgn0262872	milt	FBgn0051957	CG31957	FBgn0034300	CG5098
FBgn0053977	CG33977	FBgn0036402	CG6650	FBgn0086680	vvl	FBgn0003227	rec
FBgn0030217	CG2124	FBgn0029157	ssh	FBgn0051158	Efa6	FBgn0266580	Gp210
FBgn0024194	rasp	FBgn0263006	SERCA	FBgn0039858	CycG	FBgn0001624	dlg1
FBgn0044510	mRpS5	FBgn0284220	Top2	FBgn0003638	su(w[a])	FBgn0264078	Flo2
FBgn0265623	Su(z)2	FBgn0032586	Tpr2	FBgn0036577	CG13073	FBgn0035049	Mmp1
FBgn0015773	NetA	FBgn0036249	CG11560	FBgn0284256	bsf	FBgn0053531	Ddr
FBgn0004854	B-H2	FBgn0037890	CG17734	FBgn0032600	BuGZ	FBgn0003206	Ras64B
FBgn0014133	bif	FBgn0037439	CG10286	FBgn0039827	CG1544	FBgn0038146	CG9799
FBgn0030269	CDK2AP1	FBgn0026313	X11L	FBgn0036825	RpL26	FBgn0263855	BubR1
FBgn0262738	norpA	FBgn0039528	dsd	FBgn0039600	CG1646	FBgn0038100	Paip2

FBgn0030240	CG2202	FBgn0027556	CG4928	FBgn0022985	qkr58E-2	FBgn0037336	CG2519
FBgn0028380	fal	FBgn0028579	phtf	FBgn0011584	Trp1	FBgn0052204	CG32204

5. Chapter III - Regulatory Divergence in the *Drosophila melanogaster* subgroup

The manuscript '**Regulatory Divergence in the *Drosophila melanogaster* subgroup**' is the result of a side project that is initially based on allele specific expression analysis (ASE) started by Dr. Torres-Oliva, M.

My contributions for this manuscript includes the following parts:

- Conceptualization of the project (together with Dr. Nico Ponsien)
- Bioinformatics analyses (ATAC-seq)
- Data interpretation (together with Dr. Nico Posnien)
- Writing of manuscript draft
- Visualization

Contribution of other authors includes:

- Dr. Torres-Oliva, M. and Dr. Almudi, I. generated the transcriptomic and ATAC-seq dataset. The Illumina sequencing was performed at the Transcriptome Analysis Lab (TAL) in Göttingen.
- Data prepared by Dr. Torres-Oliva, M. include the differential expression (ASE) analysis of the parental and hybrid RNA-seq datasets and the allele specific expression analysis (ASE), presented in Figure 1B. She also kindly provided the tables with all genes and the corresponding divergence types, which underlie GO analysis, presented in Figure 1C.

Status of the manuscript:

In preparation for submission

Title

Regulatory divergence in closely related *Drosophila* species depends on the architecture of developmental gene regulatory networks.

Authors

Torres-Oliva, M.[2+]; Buchberger, E.[1+]; Almudi, I.[3]; Lu, T-H.[1], Posnien, N.[1]* (in prep.):

[1] Universität Göttingen, Department of Developmental Biology, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

[2] Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, University Hospital Schleswig-Holstein, Kiel, Germany

[3] CABD (CSIC/UPO/JA), DMC2 Unit, Pablo de Olavide University campus, Ctra. Utrera km1, 41013 Seville, Spain

[+] equal contribution

* author for correspondence: nposnie@gwdg.de

5.1. Introduction

In recent years it became clear that differences in gene expression levels contribute to a large extend to the diverse morphology of body plans that we can observe in the animal kingdom (e.g. (Carroll, 2005; Khaitovich et al., 2006; King and Wilson, 1975; Tautz, 2000)). The process of transcription i.e. gene expression must be tightly controlled to ensure the proper development and function of tissues and organs of an organism. Today, we have a very detailed understanding of transcriptional regulation that is at play on different levels. For instance, the interplay of transcription factors with co-factors and the accessibility of *cis*-regulatory regions, where these factors can bind to, represent central regulatory mechanisms (reviewed in (Buchberger et al., 2019)). But also, higher order chromatin structure (reviewed e.g. (Furlong and Levine, 2018) and post-translational processes, such as the action of regulatory RNA molecules (Bartel, 2018; Kittelmann and McGregor, 2019) work together to ensure time and tissue specific gene expression. Since evolutionary changes on each of these levels could cause natural variation in gene expression levels, the complexity of regulatory mechanisms contributes to the complications to pinpoint the exact cause of gene expression divergence between individuals and species (reviewed in (Buchberger et al., 2019)). The dissection of the molecular basis of gene expression differences and its impact on morphological evolution is further hampered by the polygenic nature of many phenotypes (Boyle et al., 2017; Mackay et al., 2009).

The number of differentially expressed genes is highly correlated with the phylogenetic distance of populations or species (Khaitovich et al., 2006; Rifkin et al., 2003). Variation in gene expression results from mutations in the genome, which can either affect the regulatory region of the differentially expressed gene itself (i.e. *cis*-regulatory changes) or an upstream regulator of the gene (for instance a transcription factor) (i.e. *trans*-regulatory changes)(Wittkopp, 2005). Even though it is under debate, how much of the overall gene expression divergence is caused by *cis*- or *trans*-regulatory changes, it is clear that both contribute to evolutionary changes within populations, strains or species (Genissel et al., 2007; Hoekstra and Coyne, 2007; Stern and Orgogozo, 2008; Tautz, 2000). The general idea is though that *cis*-regulatory changes are the main cause for divergent expression, since a change in these parts of the locus would only affect the respective gene, whereas an upstream *trans*-regulatory change would have major pleiotropic effects on its many target genes (Carroll, 2005; Prud'homme et al., 2007; Wray,

2007, 2003). Interestingly, recent studies in various model organisms revealed that gene expression changes in *cis* and *trans* are context dependent, i.e. the respective tissue and environmental factors can affect the relative contribution of each divergence type (Duveau et al., 2017; Reuveni et al., 2018).

Since mutations affecting gene expression are heritable, allele specific expression analyses (ASE) have been used to understand the molecular mechanisms of divergent gene expression. This approach takes advantage of the possibility to obtain viable offspring from crosses among closely related strains or species (Cowles et al., 2002; Wittkopp et al., 2004). Comparing the gene expression levels in closely related parental strains to the expression of the respective allele in their offspring, allows to classify the modes of differential gene expression in the parental into *cis*- and *trans*-regulatory changes. ASE allows to deduce the mode of expression changes for each expressed gene in parental species but cannot provide information about the detailed genetic causes that underlie the observed expression divergence. The combination of open chromatin datasets revealing potential regulatory regions and RNA-seq datasets, helps nowadays to better draw the links between changes in these *cis*-regulatory regions and the resulting differences in gene expression (e.g. (Hughes et al., 2017; Rendeiro et al., 2016; Starks et al., 2019)). Nevertheless, a genome-wide understanding of how *cis*- and *trans*-regulatory changes can be recapitulated on the level of accessible chromatin regions is missing up to now.

To address this open question, we use three species, *D. melanogaster, D. mauritiana* and *D. simulans* to understand the evolution of gene expression divergence during head and eye development in the *Drosophila melanogaster* subgroup. The three species vary remarkably in their eye sizes and head shapes. Differences in eye size between *D. melanogaster* and *D. mauritiana* are mainly due to variation in ommatidia number, and differences in eye size between *D. mauritiana* and *D. simulans* result mainly from variation in ommatidia size (Posnien et al., 2012). Previous gene-expression studies using these three species have shown, that a plethora of genes is differentially expressed between *D. melanogaster* and *D. mauritiana* (Buchberger et al. in prep.) and *D. mauritiana* and *D. simulans* (Almudi et al. in prep.). However, the underlying mechanisms of this gene expression divergence are completely unknown

We therefore performed ASE using the F1 hybrid generation of *D. melanogaster* x *D. mauritiana* and D*. mauritiana* x *D. simulans* to analyse the contribution of *cis*- and *trans*-

regulatory changes. We further used a comparative ATAC-seq dataset to address the question how changes in the regulatory landscapes influence species-specific gene expression. We assessed and compared open chromatin regions between *D. melanogaster*, *D. mauritiana* and *D. simulans* in terms of differential *cis*-regulatory landscapes and sequence divergence and revealed that genes that were found to be differentially expressed due to *cis*-regulatory changes indeed exhibited a more divergent chromatin architecture. Additionally, orthologous regulatory sequences of *cis*-divergent genes showed a more pronounced sequence variation than regulatory regions of conserved genes or genes that are differentially expressed due to *trans*-changes. We suggest that both mechanisms contribute to *cis*-regulatory changes – namely differential accessibility of regulatory regions, but also sequence divergence in potential promoters or enhancers.

5.2. Results

5.2.1. Regulatory Divergence in the *D. melanogaster* subgroup

To better understand the patterns of regulatory divergence in closely related *Drosophila* species, we obtained F1 hybrids of interspecific crosses between *D. simulans* and *D. mauritiana* and between *D. melanogaster* and *D. mauritiana*. The three species vary extensively in their adult head shapes ((Posnien et al., 2012), Figure 21A) and therefore provide an excellent model to understand how conserved GRNs evolve over time, being on the one hand tightly controlled to ensure proper organ development and function, and on the other hand flexible enough to allow evolution of size and shape of these structures. Since the adult head structures in *Drosophila* develop from two 2D larval epithelia, the eye-antennal discs (Haynie and Bryant, 1986) and variation in adult morphologies arise from differences during development of the respective structures, we applied RNA-seq to mid L3 developing eye-antennal discs (96h after egg laying, (AEL)) of the three parental species and the two hybrid crosses (Figure 21A). We first determined the number of genes, differentially expressed between the parental species. Subsequently, we analysed differential expression of the parental alleles in the two hybrid datasets to assess the type of regulatory divergence for each gene. When a gene was differentially expressed in the parental species as well as their alleles in the hybrids, its expression diverged due to *cis*-regulatory changes in its own locus. *Cis*-regulatory changes can have two underlying causes. First, mutations in either promoter or enhancer regions can change the regulation of a gene via, for instance, affecting the binding of transcription factors

(e.g. (Prud'homme et al., 2006; Rogers et al., 2013)). Second, variation in the accessibility of such regulatory regions could alter gene regulation and hence gene expression (e.g. (Zhang and Borevitz, 2009)). If we found a gene to be differentially expressed between the parentals, but the two alleles in the hybrids were not, the gene expression in the parental species diverged due to upstream *trans*-regulatory changes. *Trans*-regulatory changes can arise due to nucleotide changes in the coding sequences (CDS) of an upstream regulator, but also due to *cis*-regulatory changes of an upstream transcription factor which changes the amount of available upstream regulators (Wittkopp, 2005). Genes that show neither significant differential expression between the parentals nor in the hybrid setting are considered conserved in expression. If genes with conserved expression levels in the parental species, showed differential allelic expression in the hybrids, we called the regulation type 'compensatory'.

Comparing *D. melanogaster* and *D. mauritiana*, we found that most genes were conserved (71%, Figure 21B). Of the genes that showed divergent expression, most were differentially expressed due to variation in *trans*, i.e. due to variation in an upstream regulator (20%, Figure 21B). Only 5% of the differentially expressed genes showed regulatory divergence in *cis* (5% Figure 21B). A nearly equal number of genes showed signatures of compensatory divergence (4%, Figure 21B). Although the general trends are the same for both pairwise comparisons, we found more genes to be differentially regulated by *cis*-regulatory effects between *D. melanogaster* and *D. mauritiana* (Figure 21B), compared to the more closely related *D. simulans* and *D. mauritiana* (2% *vs.* 5%, Figure 21B). In the latter comparison we found even more genes to be differentially expressed due to *trans*-regulatory divergence (23% *vs.* 20%, Figure 21B).

We further assessed in which biological processes these genes were enriched and performed a GO enrichment analysis. Genes which are conserved between *D. melanogaster* and *D. mauritiana* were involved in basic processes like cell morphogenesis, cell proliferation, growth and developmental processes (Figure 21C). Genes that show regulatory divergence in *trans*, were enriched in morphological and developmental processes, and more specifically in transcription, neuronal processes or photoreceptor development (Figure 21C). We found genes that were differentially expressed due to *cis*-regulatory divergence to be enriched in metabolic and biosynthetic GO terms (Figure 21C). Genes with signs of compensatory regulation were mainly enriched in GO terms like cell fate, cell cycle or larval cuticle development (Figure 21C).

We obtained a very similar pattern for the other species pair (*D. simulans vs. D. mauritiana*) (Supplementary Figure 21). Overall, we found that most genes that were differentially expressed in the developing eye-antennal disc in closely related species were different due to *trans*-regulatory changes. We showed that genes diverging in *cis* or *trans* take part in different processes, with the latter ones being enriched in developmental GO terms.

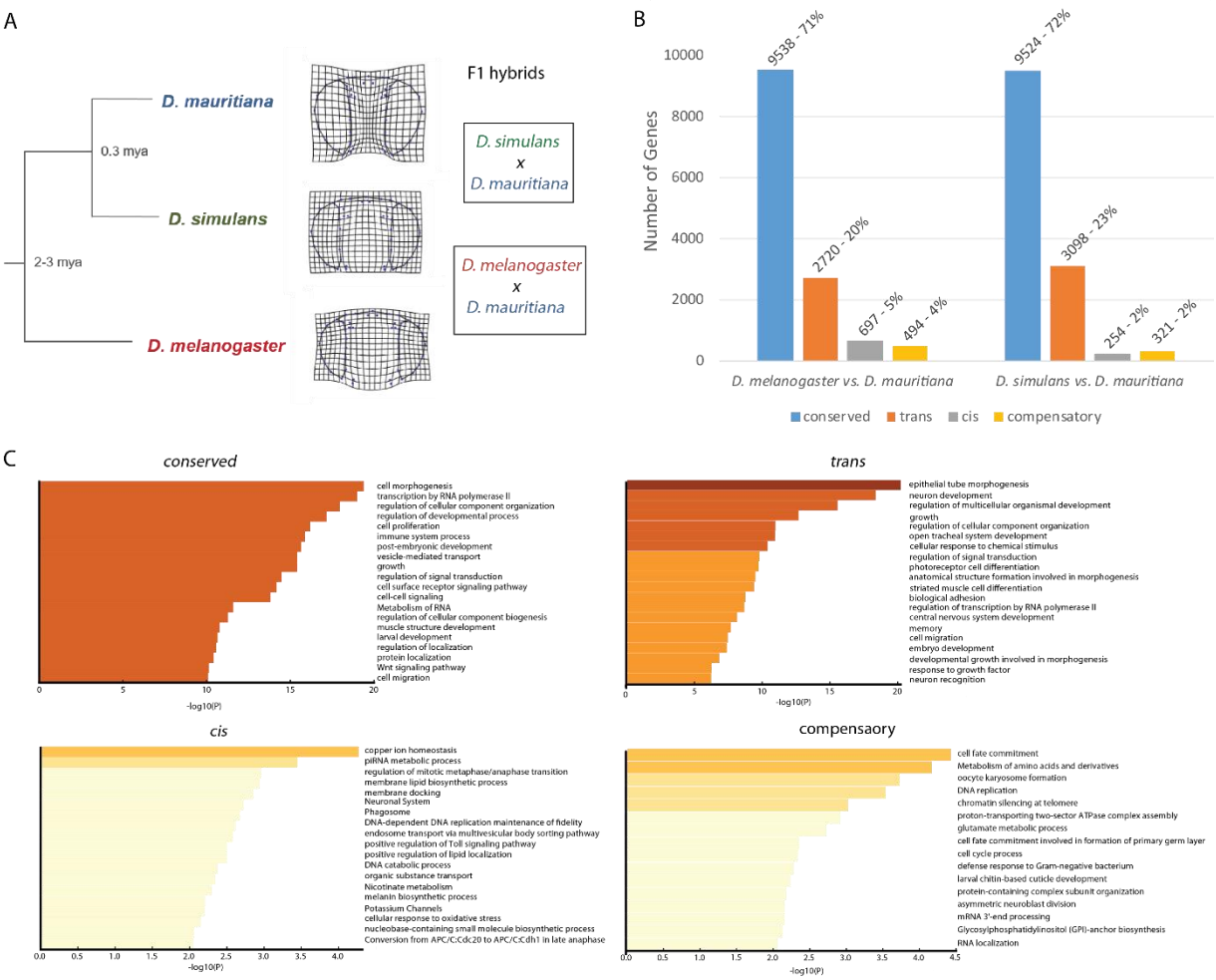


Figure 21. Regulatory divergence in three closely related *Drosophila* species. A. Phylogenetic relationship between the three *Drosophila* species used in this study and differences in their head shapes (Posnien et al., 2012). The boxes summarize the two crosses giving rise to viable F1 hybrids. RNA-seq was performed of developing eye-antennal discs of the three parental species and the two F1 hybrid offspring. **B.** Allele specific expression analysis between the two species pairs (*D. melanogaster vs. D. mauritiana* and *D. simulans vs. D. mauritiana*). Most genes are conserved between the species. Differentially expressed genes are predominantly differentially expressed due to changes in *trans*. We found more '*trans*-genes' between *D. simulans vs. D. mauritiana.* A higher number of *cis*-regulatory changes was observed between *D. melanogaster vs. D. mauritiana*. **C.** A random subset of 3000 conserved genes (see Material and Methods) between *D. melanogaster* and *D. mauritiana* was highly enriched in processes like morphogenesis, cell proliferation, growth, larval development etc. Genes that were differentially expressed due to *trans*-regulatory changes were as well highly enriched in developmental and morphogenetic processes (especially neuronal development), whereas we found more biosynthetic and metabolic GO terms for *cis*-effect genes. Genes with compensatory expression in hybrids were mainly enriched in cell-cycle but also metabolic processes.

5.2.2. A comparative ATAC-seq dataset of three closely related *Drosophila* species

To test if the different types of regulatory divergence during head development can also be recapitulated on the level of the regulatory landscape, we generated a comparative ATAC-seq dataset for eye-antennal discs (96h AEL) for the three species. We could significantly call 21,705 peaks in *D. melanogaster*, 21,499 peaks in *D. mauritiana* and 20,374 peaks in *D. simulans* (Figure 22A). These numbers of open chromatin regions in the developing eye-antennal disc is in concordance with previous studies (Davie et al., 2015).

For further quality assessment of the three ATAC-seq datasets, we calculated the insert size distribution. This is based on the assumption that the Tn5 transposase used for ATAC-seq can only insert adapters where the DNA is not covered by nucleosomes. Therefore, proper ATAC-seq library preparation should result in a clear peak at ~100bp where the DNA is depleted of nucleosomes, hence most easily accessible to the transposase, and smaller peaks resulting from sequences that are wrapped around different-sized nucleosomes. All of our datasets showed the typical periodicity of ~200 bp (Figure 22B, Supplementary Figure 22, (Buenrostro et al., 2013; Davie et al., 2015)). We annotated open chromatin regions of *D. melanogaster* to the gene loci according to the closest transcription start site (TSS) and to the respective gene features. We found that open chromatin sites predominantly mapped to promoter regions (~37%), intronic (~30%) and intergenic regions (~18%) (Figure 22B). This demonstrates that we generated a reliable open chromatin dataset for developing eye-antennal discs at mid L3 larval stages in terms of peak number and annotation.

A

Species	original # of Peaks	# of Peaks afterCoordinate Conversion	# of unmapped Peaks after Coordinate Conversion
D.melanogaster	21,705		
D. mauritiana	21,499	20,678 (96%)	821 (4%)
D. simulans	20,374	19,732 (97%)	642 (3%)

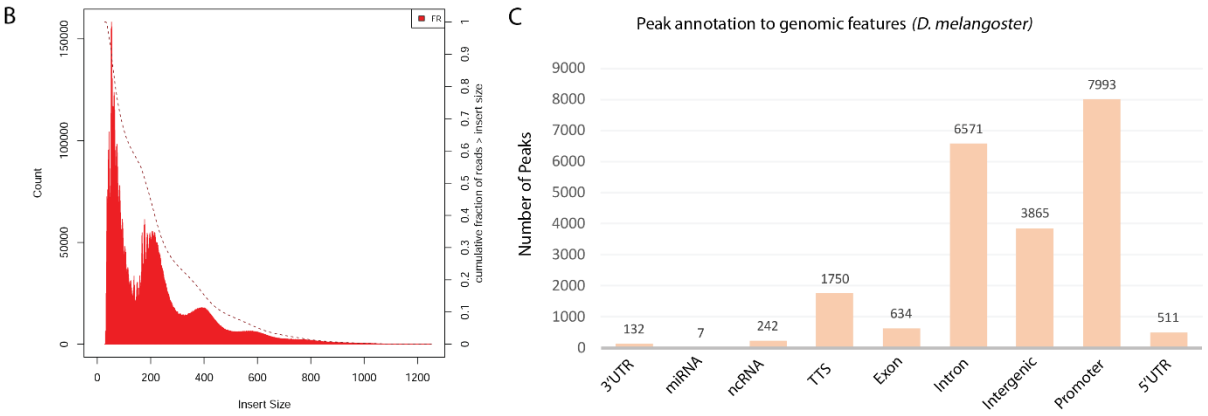


Figure 22. A comparative ATAC-seq dataset for the *Drosophila melangaster* subgroup. A. We were able to call a similar number of open chromatin peaks in all three species, namely between 20,300 and 21,700, which is comparable with previous ATAC-seq studies (Davie et al., 2015). We converted the peak coordinates of *D. mauritiana* and *D. simulans* into the *D. melanogaster* coordinate system. Using our customized pipeline, we were able to convert 96-97% of all peaks in *D. mauritiana* and *D. simulans*, respectively. **B.** The insert sizes of the *D. melanogaster* ATAC-seq dataset shows the typical periodicity of the expected ~200 bp (Buenrostro et al., 2013). **C.** Annotation of the *D. melanogaster* peaks to gene features. Typically for open chromatin datasets, most peaks mapped to either promoter, intronic or intergenic regions.

To compare the open regulatory landscape of *D. melanogaster* with the ones of its sister species, we developed a pipeline to convert *D. mauritiana* and *D. simulans* peak coordinates into the *D. melanogaster* genome coordinate system. For this we adapted the workflow used by the UCSC coordinate conversion tool (http://genome.ucsc.edu/, see Material and Methods and Appendix). In short, split chromosomes of *D. melanogaster* were aligned to the *D. mauritiana* or the *D. simulans* genome and respective chain files were generated for each of the two species. We then used the liftOver tool (Hinrichs et al., 2006) to convert the coordinates. By this we were able to convert 96% and 97% of *D. mauritiana* and *D. simulans* peak regions, respectively into *D. melanogaster* coordinates (Figure 22A). Peaks that could not be reliably converted were mostly found at the centromeres of chromosomes (Supplementary Figure 23A), which is consistent with the suggested quick evolution of these genomic regions (Henikoff et al., 2001). To overcome this bias, we removed the centromeric regions in the genome of *D. melanogaster* for further analyses (Supplementary Figure 23B, Material and Methods). Annotation of the converted open chromatin regions to gene features showed that they mainly mapped to intronic, intergenic and promoter regions which is comparable to the annotation of *D. melanogaster* ATAC-seq peaks (Figure 22C, Supplementary Figure 23C and D).

After applying this pipeline for peak coordinate conversion, we continued with a total number of 20,678 peaks in *D. mauritiana* and 19,732 peaks in *D. simulans*, compared to 21,705 peaks in *D. melanogaster*.

5.2.3. Genes with species specific regulatory regions are more often regulated in *cis*

To understand how the accessibility of regulatory regions influences the evolution of gene expression in the *D. melanogaster* subgroup, we sought to compare the open-chromatin landscape between each of the species pairs *(D. melanogaster vs. D. mauritiana* and *D. simulans vs. D. mauritiana)*. For each species we first summarized the peaks that mapped either to a TSS/promoter region or an intronic region and excluded peaks that mapped to intergenic regions (also see Technical and other considerations). To find orthologous peak regions between *D. melanogaster* and *D. mauritiana* or *D. simulans* and *D. mauritiana*, we overlapped peak regions using the bedtools suite (Quinlan, 2014; Quinlan and Hall, 2010), which resulted in three peak sets per pairwise comparison: 1) Peaks shared between two species; 2) Species-specific peaks for one species; and 3) Species-specific peaks for the other species. We found 11,439 peaks, mapping to 6,159 genes that were shared between *D. melanogaster* and *D. mauritiana*, and 3,103 and 2,829 peaks being specific for *D. melanogaster* and *D. mauritiana*, respectively. Of these, a higher number of species-specific peaks are annotated to introns than promoters (Table 5). We found a very similar pattern between *D. mauritiana* and *D. simulans*, but a slightly smaller number of species-specific peaks between these two closer related species (Table 5).

Table 5. For each species in each species comparison we combined intronic and promoter/TSS peaks (excluding intergenic peaks) and overlapped the peak sets to find orthologous peaks, and species-specific peak sets. The table lists intronic peaks in each species, plus the overlapping intronic peaks in the fist column, and the same information for TSS/propmter peaks in the second column. The sum of species specific intronic, TSS/promoter and shared peaks is shown in the thrid column (please note, that this number does not correspond to the total peak number per species after peak conversion). In all comparisons we found more species-specific peaks mapping to intronic regions, than TSS/promoter regions. The % is calculated by deviding the species specific intronic or TSS/promoter peaks by the sum of intronic or TSS/promoter peaks (species-specific plus shared).

	intronic peaks	**TSS/promoter peaks**	**sum of TSS/promoter and intronic peaks / species (species specific + shared)**
	D. melangoaster vs. D. mauritiana		
D. melanogaster	1696 (12.2%)	1407 (10.1%)	13898
D. mauritiana	1718 (12.6%)	1111 (8.2%)	13624
shared	4488	6307	
	D. simulans vs. D. mauritiana		
D. simulans	1021 (8%)	854 (6.7%)	12755
D. mauritiana	1537 (11.3%)	1223 (9%)	13640
shared	4707	6173	

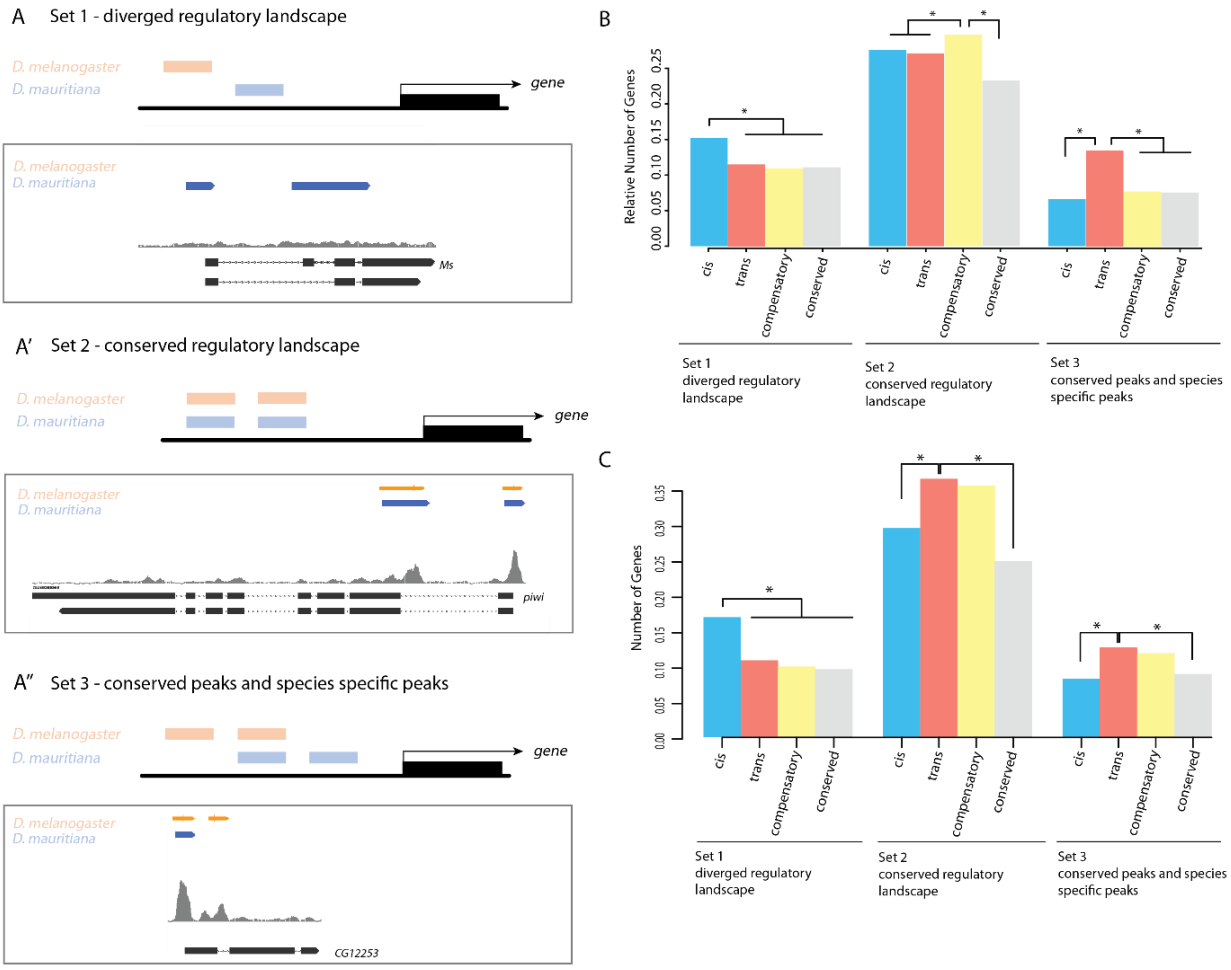


Figure 23. Comparison of the regulatory landscape. A-A''. We summarized all gene loci according to their regulatory landscape into **A)** genes with highly diverged regulatroy landscape, with no overlapping peaks between the species. **A')** genes with a very conserved regulatory landscape and no species-specific peaks and **A'')** genes with overlapping peaks but additional species specific ones. The loci in the rectangles show one randomly picked locus from each Set with the respective ATAC-seq peaks in the two species. orange: *D. melanogaster*, blue: *D. mauritiana*, grey: read densitiy of the *D. melanogaster* ATAC-seq dataset. **B-C.** Genes with a highly divergent regulatory landscape are significantly more often differentially expressed due to *cis*-regulatory changes. We find a high number of compensatory changes for genes with a very conserved regulatory landscape and significantly more genes that are differentially regulated in *trans* for genes with overlapping but also species-specific genes. Note that we provide all p-values between the pairwise comparisons in Supplementary Table 20A and B. **B.** *D. melanogaster vs. D. mauritiana* **C.** *D. simulans vs. D. mauritiana.*

Second, we assigned every gene locus to one of three genes sets (Figure 23A-A''): The first gene set included genes that showed a completely divergent regulatory landscape (i.e. no overlapping peaks between two species) (Figure 23A). The second set included genes, that had the same regulatory landscape in two species (i.e. only overlapping peaks between two species, Figure 23A') and third we pooled genes that had a similar open chromatin landscape in both species, but also putative species-specific regulatory regions (Figure 23A''). We then overlapped these sets with the differentially expressed genes, for which the type of regulatory divergence was known (Figure 21B). Interestingly, we found that genes that are differentially

regulated due to *cis*-regulatory changes between the *D. melanogaster* and *D. mauritiana*, overlap predominantly with genes in Set1, i.e. genes that have only species-specific peaks (Figure 23B and C , e.g. *cis vs. trans* Fishers exact test, p=0.0095), whereas genes with a conserved regulatory region, are mainly differentially expressed between species, due to compensatory mechanisms and surprisingly not necessarily conserved in expression levels (Figure 23B). Genes, which show conserved regulatory regions but also additional species-specific peaks (Set3), were predominantly differentially expressed due to *trans* mechanisms (Figure 23B).

To test, whether we find similar patterns in more closely related species, we also performed this analysis for the set of differentially expressed genes between *D. mauritiana* and *D. simulans*. As observed for *D. melanogaster* and *D. mauritiana* genes in Set 1, summarizing genes with a divergent chromatin landscape, were mostly differentially regulated due to variation in *cis*. However, for genes with a conserved regulatory landscape we found that they were differentially expressed due to *trans*-regulatory changes. Genes in Set 3 showed the same pattern with genes being mostly differentially regulated due to upstream *trans* mechanisms (Figure 23C).

Overall, we show that genes with a highly divergent DNA accessibility landscape were significantly more often differentially expressed due to *cis*-regulatory changes, compared to genes that show a more conserved regulatory architecture.

5.2.4. Regulatory regions of genes, diverging in cis, show a higher sequence divergence

Since *cis*-regulatory divergence may not only arise due to differences in accessibility of the respective regulatory regions, but also due to sequence changes affecting for instance transcription factor binding, we focused in more detail on the sequence divergence of orthologous open chromatin regions between the species. We extracted the sequences of all orthologous intronic and TSS peaks and compared their sequences between the species pairs. Peak regions that were annotated to genes showing *cis*-regulatory divergence, showed a significantly lower percentage of sequence identity between *D. melanogaster* and *D. mauritiana*. Peaks assigned to genes showing compensatory divergence had similarly diverged regulatory sequences (Figure 24A). We found the same trend between *D. mauritiana* and *D. simulans*, although the differences in sequence divergence between different regulatory

groups were not significant, reflecting the closer phylogenetic relationship between these two species (Supplementary Figure 24A).

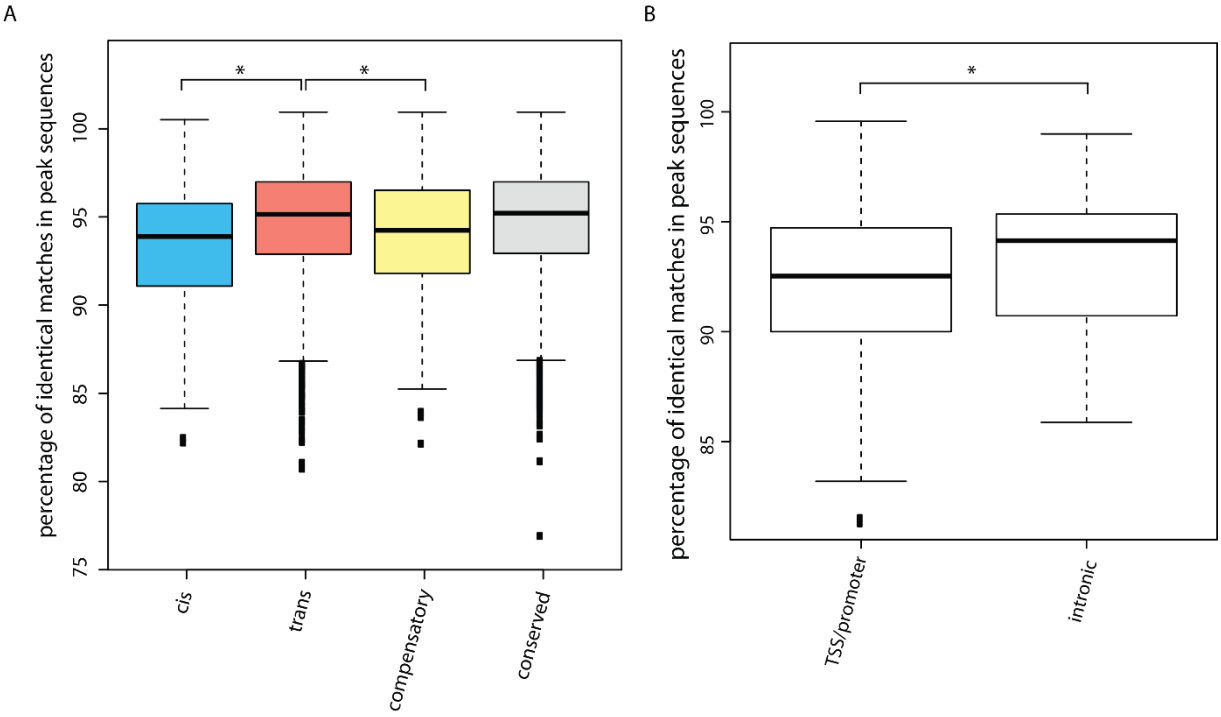


Figure 24. *Cis*-regulatory changes arise due to differences in DNA accesssibility but also due to sequence divergence. A. Genes that are differentially expressed due to *cis*-regulatroy changes, but also compensatory genes show more diverged peak sequences than conserved genes or genes with *trans* effects. Note that we provide all p-values between the pairwise comparisons in Supplementary Table 21. **B.** We separated TSS/promoter peaks from intronic peaks and showed that intronic peak sequences are on average more conserved than peaks mapping to promoter regions. Note that we provide all p-values between the pairwise comparisons in Supplementary Table 22.

We further wanted to test, whether TSS/ promoter regions and intronic regions evolve quicker in terms of nucleotide content. For this, we aligned the sequences of accessible orthologous promoter regions and intronic regulatory regions of *D. melanogaster* and *D. mauritiana*. Interestingly, we observed for all four classes of genes (*cis*, *trans*, compensatory and conserved) that intronic sequences seem to be more conserved in terms of nucleotide sequences, whereas peaks in the promoter regions show a lower sequence similarity (Figure 24B, Supplementary Figure 25, Supplementary Table 22).

We conclude, that orthologous regulatory sequences of genes, differentially expressed due to *cis*-regulatory changes show higher sequence divergence and that this pattern is more pronounced in peaks annotated to TSS/promoter regions.

5.2.5. Regulatory divergence in transcription factors

Since the overall patterns seemed to be similar between the two species comparisons, we next asked, whether the same gene sets were composed of the same genes in the two comparisons. For this we focused on a set of 149 transcription factors which we overlapped with the gene sets shown in Figure 23A-A''. Furthermore, we assessed if these transcription factors, if differentially expressed, showed divergence in *cis*, *trans* or compensatory expression in the hybrids. We found strikingly few transcription factors in Set 1 (i.e. genes with only species-specific peaks, Supplementary Table 23). The potential regulatory regions of most transcription factors were conserved (Set 2, 72 and 79 TFs from the *D. melanogaster* vs. *D. mauritiana* and *D. simulans* vs. *D. mauritiana* comparison, respectively; Supplementary Table 23) or only slightly diverged with additional species-specific peaks (Set 3, 51 TFs from the *D. melanogaster* vs. *D. mauritiana* and 46 TFs from *D. simulans* vs. *D. mauritiana* comparison, respectively; Supplementary Table 23). Consistently, we found transcription factors mostly to be conserved in expression, and if differentially expressed, this was due to *trans*-regulatory changes, except in one case, where we found *bicoid* (*bcd*) to be differentially regulated between *D. melanogaster* and *D. mauritiana* due to a *cis*-regulatory change.

Overall, we found that the expression and regulatory landscape of transcription factors is highly conserved between the two species pairs.

5.3. Discussion

5.3.1. Regulatory divergence is context dependent

Our differential expression analysis between three closely related *Drosophila* species revealed that most genes were conserved among species. This recapitulates previous data which showed that the overall gene expression dynamics in developing eye-antennal discs between the three species *D. melanogaster*, *D. mauritiana* and *D. simulans* are to a large extend conserved (Torres-Oliva, 2016). Still, we found a substantial number of genes to be differentially expressed during eye and head development (Buchberger et al. in prep; Chapter II, Almudi et al. in prep). Since changes in gene expression during development often correlate with variation in adult morphology and physiology (Carroll, 2005; Khaitovich et al., 2006; King and Wilson, 1975; Tautz, 2000), this observation most probably recapitulates the remarkable variation in size and shape of the head cuticle and the adjacent compound eyes (Posnien et al., 2012). However, in most cases it is not known which regulatory change underlies differential

expression of a gene. ASE analysis provides a powerful tool to test this on a genome-wide level. Therefore, we generated F1 hybrids between *D. melanogaster* x *D. mauritiana* and *D. simulans* x *D. mauritiana* and performed ASE analysis do understand if the cause for differential expression in a developing epithelium can be mainly found in the gene's own regulatory region (*cis*) or is rather caused by changes in upstream factors (*trans*).

Interestingly, the majority of the differentially expressed genes was due to variation in *trans.* We applied the same analysis pipeline to two species comparisons differing in their divergence time using *D. melanogaster vs. D. mauritiana*, which diverged about 2-3 Mya, and *D. simulans vs. D. mauritiana* diverging about 0.3 Mya (Figure 21A). Even though patterns in terms of regulatory divergence were similar, the number of *cis*-regulatory changes increased with phylogenetic distance, whereas the number of *trans*-regulatory changes decreased (Figure 21B). This is consistent with the finding that usually more *cis*-regulatory changes accumulate throughout time (Metzger et al., 2017; Stern and Orgogozo, 2008; Wittkopp et al., 2008).

The excess of *trans*-regulatory changes contradicts most previous studies which reported a higher contribution of *cis*-regulatory changes, compared to *trans* (e.g. (Graze et al., 2009; Wittkopp et al., 2008, 2004). This has mainly been explained by the fact that a *cis*-regulatory change would only affect the respective locus, whereas changes in an upstream regulator would have more widespread and pleiotropic effects on all of its target genes (Wittkopp et al., 2008). Nevertheless, other studies also found a slightly higher amount of *trans*-regulatory changes (McManus et al., 2010; Suvorov et al., 2013). More *trans*-acting changes were mostly found in intraspecific comparisons, explained by a larger mutational target size of *trans*-factors that correlates positively with mutational variance (i.e. increase in trait-variance introduced by mutations in each generation (Landry et al., 2007)) (Landry et al., 2007; Wittkopp et al., 2008). McManus and colleagues also found an increase in *trans*-regulatory changes between *D. melanogaster* and *D. sechellia* and explained that pattern with the small population size of the latter species. Consequently, mutations would rather get fixed due to random genetic drift than due to natural selection (McManus et al., 2010). We cannot exclude a similar scenario for *D. mauritiana*, which is endemic on the island of Mauritius (David et al., 1989). To test this, it would be necessary to produce an F1 hybrid generation between the two cosmopolitan species *D. melanogaster* and *D. simulans* and compare the ASE analysis for this comparison with the already existing ones, including *D. mauritiana*.

We examined the role of the differentially expressed genes in the four divergence types to test, whether there may be functional constraints on the type of regulatory divergence. Indeed, we found that genes differentially expressed due to variation in *trans* were enriched in developmental processes. In contrast, genes showing *cis*-regulatory divergence were in general enriched in metabolic and biosynthetic processes. A similar pattern has been observed in *Drosophila* embryos, where genes with *cis*-effects were more enriched in housekeeping functions, and genes with *trans*-effects mainly functioned in developmental and gene regulatory processes (Cannavò et al., 2017), suggesting that the function of a gene product indeed has an impact on its evolvability.

Another source of constraints may be imposed by the excess of regulatory interactions of a gene within a gene regulatory network (GRN). We studied a developing tissue, whereas comparable ASE studies were mostly performed in whole-body adult flies (McManus et al., 2010; Suvorov et al., 2013). We therefore checked specifically for regulatory divergence of transcription factors to find out if upstream, developmental regulators are more constraint and as previously suggested, more likely to be affected by *trans*-regulatory changes (Luscombe et al., 2004; Wittkopp, 2005). Likewise, genes showing *cis*-regulatory changes display a lower average connectivity in mouse tissues (Mack et al., 2019) but also in plants (Mähler et al., 2017). In general, we found a low number of TFs to be differentially expressed between the species. If they showed divergent expression, this was due to upstream *trans*-regulatory changes, which suggests, that the loci of these important regulators are indeed kept highly conserved between the species. We did not specifically analyse the connectivity of these TFs but one can assume that most of developmentally important TFs are positioned at the top of the GRNs and are most probably highly interconnected (MacNeil and Walhout, 2011; Stern and Orgogozo, 2008). Therefore, an excess of *trans*-regulatory divergence may be a common feature of developing tissues.

Nevertheless, several important regulators in the eye-antennal disc were differentially expressed due to differences in an upstream regulator. Between *D. melanogaster* and *D. mauritiana* these include for instance *sine oculis* (*so*), *pannier* (*pnr*), *ocelliless* (*oc*), whereas we found that the alleles of *eyeless* (*ey*) were only differentially regulated in the hybrid (Supplementary Table 23). We could functionally show, that differential expression of *pnr* indeed underlies differences in head shape and eye size between these two species

(Buchberger et al. in prep.; Chapter II). Finding the potential upstream *trans*-acting transcription factors, that lead to differential expression of these important developmental regulators might eventually reveal the true genetic causes of variation in adult morphologies that we observe between the here studied *Drosophila* species. Also, between *D. simulans* and *D. mauritiana* transcription factors involved directly in eye development were differentially expressed. Among these are interesting candidates like the proneural gene *atonal* (*ato*), *scalloped* (*sd*), or *ttk* (*tramtrack*), for which a role in ommatidia development was reported (e.g. (Garg et al., 2007; Jarman et al., 1994; Li et al., 1997; Siddall et al., 2009)). In the light of the finding, that eye size between these two species varies due to changes in ommatidia size (Posnien et al., 2012), these could be additional candidates to test if they indeed impact the size of the individual facets.

5.3.2. *cis*- regulatory divergence is due to changes in chromatin accessibility and sequence divergence

We further tested, which mechanisms contribute to *cis*-regulatory divergence in our data. Two reasons can theoretically underlie *cis*-regulatory changes leading to subsequent gene expression divergence, namely either mutations directly in the regulatory regions or divergent accessibility of these regions.

Orthologous regulatory sequences might have experienced changes in their nucleotide sequence, which could, amongst other things, affect TF-binding (Wittkopp, 2013). Even if the regulatory regions of a gene are characterized, studying the influence of sequence changes on gene expression is not straightforward. In some reported cases only one nucleotide change is enough to alter the temporal expression of an important master regulator (Ramaekers et al., 2018), whereas other enhancer sequences keep their conserved function despite extensive reshuffling of TF binding sites (Khoueiry et al., 2017; Ludwig et al., 2000). However, some mechanistic insights have been gained in the last years, that may help interpreting the obtained data. It was for instance shown in *Drosophila*, that quantitative changes in enhancer strengths between species correlate linearly with sequence divergence (Arnold et al., 2014) and that sequence changes in regulatory regions may lead to differential functionality due to loss in transcription factor or co-factor binding (e.g. (Paris et al., 2013; Schmidt et al., 2010; Zheng et al., 2010)). However, how deleterious the loss of a certain TF binding motif is, seems to depend on the combinatorial binding of a TF collective (Khoueiry et al., 2017). In our genome wide comparison, orthologous sequence divergence is higher in open chromatin regions close to

genes with *cis*-regulatory divergence between species or compensatory changes in hybrids. A higher rate of polymorphisms in promoter regions of *cis*-effect genes (compared to *trans*-effect genes) was for instance shown in plants (Zhang and Borevitz, 2009), but also in *Drosophila* (McManus et al., 2010). These studies and our results suggest higher purifying selection in regulatory regions of highly connected developmental genes, which we found to be more often differentially expressed due to upstream *trans*-effects or are conserved between the species.

Studies in *Arabidopsis thaliana* suggest that not only open chromatin regions with nucleotide changes, but also differentially accessible DNase hyperactive sites (DHS) are often found close to genes that show differential expression between ecotypes (Alexandre et al., 2018). Therefore, differential accessibility of regulatory regions very likely adds to expression variation in *cis.* Here, we found that indeed genes with highly divergent DNA accessibilities are significantly more often differentially expressed due to *cis*-regulatory changes. Chromatin remodelling and differential enhancer opening is prevalent during development (e.g. Bozek et al., 2019; Hughes et al., 2017; Kvon et al., 2014; McKay and Lieb, 2013; Uyehara et al., 2017), and in the last years an in-depth understanding of how 3D chromatin organization, epigenetic and histone modifications and chromatin accessibility interact has emerged (e.g. (Corrales et al., 2017; Cubeñas-Potts et al., 2017; Rennie et al., 2018b; Sexton et al., 2012)). How this though affects divergent DNA accessibility among species is still largely unclear.

We further checked for sequence divergence and accessibility of promoters and intronic regulatory regions separately. Regulatory sequences annotated to TSS and promoter regions showed a higher sequence divergence, whereas intronic regulatory sequences seemed to be more constraint. Intronic peaks were more often differentially accessible in both of our comparisons, suggesting that in general the accessibility of TSS/promoter peaks is more conserved, whereas accessibility of regulatory regions in introns seems to be more species specific. We could therefore observe the trend in which changes in DNA accessibility affect more often intronic regions, though their sequences seem to stay more conserved. Apart from the circumstance that intronic sequences are maybe more conserved due to their location in gene loci, higher sequence conservation was indeed observed in long introns, which are thought to harbour more functional elements (Haddrill et al., 2005). It will be important to compare these results with sequence divergence of more distant intergenic regulatory regions.

5.3.3. Compensation and conservation of gene expression

It was suggested, that gene expression falls largely under stabilizing selection (e.g. (Landry et al., 2005; Lemos et al., 2005)), i.e. that a certain level of gene expression has to be kept stable. The rational is, that even though mutations in regulatory sequences accumulate over time, *trans*-regulatory factors co-evolve to buffer these changes (Landry et al., 2005). We found in our analysis a high number of compensatory effects, characterized by allelic misexpression in the F1 hybrid generation. Interestingly, regulatory regions of genes show a similar sequence divergence than genes that are affected by *cis*-regulatory changes, suggesting that indeed upstream *trans*-regulatory factors co-evolved to maintain the expression levels in the parental species. We found compensatory regulation in all three gene sets, predominantly though in genes that show no divergence in peak accessibility, therefore, the main mode of *cis*-regulatory changes in these genes might be attributed to nucleotide changes. Nevertheless, compensatory changes are also found in genes with diverged accessibility of regulatory regions. One characteristic of enhancer function is that they usually work in a highly modular manner (reviewed for example in (Arnone and Davidson, 1997; Wray, 2003)). It was for instance estimated for *Drosophila* that each expressed gene is controlled by an average of four distinct enhancers (Kvon et al., 2014). This modularity allows also to control gene expression in a spatially and temporally controlled manner (reviewed for instance in (Prud'homme et al., 2007)). This has been elegantly shown in more simple traits like pigmentation patterns, in which the deletion of a 'spot enhancer' or an 'abdomen enhancer' leads to loss of wing pigmentation on a *Drosophila* wing or loss of dark abdomen coloration (Jeong et al., 2006; Prud'homme et al., 2006). Our dataset provides the opportunity to further analyse in more detail, how much of the compensatory coevolution is driven by differential combinatorial usage of such enhancer modules. Since DNA accessibility is highly dependent on the developmental stage and tissue, one can assume that this kind of compensatory regulation is in general highly context dependent and calls for a more thorough comparison with other developing tissues, like the wing disc for instance.

We found a high number of genes that show conserved expression in the parental species as well as in their F1 hybrids. These conserved genes were highly enriched in general developmental functions, like growth, proliferation or morphogenesis, which is consistent with our finding that most developmental TFs are conserved in expression between the species. Regulatory sequences of conserved genes were equally constraint in terms of sequence

divergence than genes that showed *trans*-regulatory divergence. Nevertheless, in cases that show high sequence divergence, conservation of TF binding could be attributed for example to the topology but also the function of the GRN. It has been suggested that upstream genes in highly connected GRNs show a more conserved TF occupancy (Khoueiry et al., 2017) and have therefore a higher chance to balance sequence changes in their regulatory regions. For conserved genes that show a high divergence in peak accessibility between the species, the modularity of enhancer elements, as discussed for compensatory changes, might ensure the correct level of gene expression. In contrast to genes that show compensatory changes though, these mechanisms would not lead to misexpression in the hybrids, therefore they might be less dependent on the co-evolution of upstream *trans*-regulators.

Overall, the high number of compensatory and conserved genes that do show changes in DNA accessibility or enhancer and promoter sequence reflects the high potential of compensatory mechanisms, that ensure the correct level of expression despite substantial *cis*-regulatory changes (Ludwig et al., 2000). In this study, we mainly concentrate on changes in regulatory regions and upstream transcription factors. Given the highly complex regulation of gene expression (reviewed in (Buchberger et al., 2019)) it remains to be studied how gene expression control on other levels, for instance miRNAs contribute to such compensatory mechanisms.

5.3.4. Technical and other considerations

The combination of RNA-seq and open chromatin datasets like ATAC-seq allows to deduce certain patterns in gene expression divergence and its correlation with regulatory regions. Apart from the high context dependency and compensatory mechanisms of gene expression, additional technical limitations must be considered. We assume here, that the annotation of a peak to the closest TSS does represent the true regulatory influence on the respective gene, which in *Drosophila* is often, but not always true. A systematic annotation of active enhancers in *Drosophila* revealed that about 88% are located in direct proximity of the target gene (Kvon et al., 2014). About 20% of all enhancers, were found to be located in between 4 kb distance from the respective TSS; about the same fraction of enhancers showed though a distance >100 kb (Kvon et al., 2014). The fact that we focus on only TSS/promoter and intronic peaks, might on the one hand reduce the chance to interpret intergenic peaks with ambiguous gene association, but on the other hand leads to wrong assumptions of peak

number in a gene's regulatory landscape. Therefore, it will be important to repeat the analysis including these intergenic peaks. The annotation of peaks to a certain gene is particularly error prone for peaks in regions of the genome that contain many overlapping gene models. This is mostly due to the fact that an enhancer element might control only one gene in a multi-gene locus, or several ones. These drawbacks have started to be overcome in studies using annotation-unbiased approaches that base the characterization of regulatory regions rather on parameters related to transcriptional properties (Rennie et al., 2018a). Also, our focus on TSS/promoter and intronic peaks reduces the number of genes to be included in the analysis. While our study focuses on 6100-6200 genes, an earlier estimation using the same RNA-seq dataset as a basis resulted in about 9000 genes being transcribed in the eye (Torres-Oliva, 2016). This clearly shows, that to understand the complete picture, also intergenic peaks have to be included in this analysis.

Furthermore, we define here 'differential accessibility' as a peak being significantly called or not, i.e. we did not consider the height of ATAC-peaks. The height of peaks is defined by the number of reads that map to a specific peak region. This can be influenced for example by the number of cells in a heterogenous tissue that show the specific chromatin opening or the 'accessibility' of a specific regulatory region. To address questions like variation of DNA accessibility in an epithelium like the eye-antennal disc, which gives rise to a plethora of different head structures (Haynie and Bryant, 1986), one might learn a lot by applying single sell ATAC-seq (Buenrostro et al., 2015; Cusanovich et al., 2015).

5.4. Conclusion

In summary, we show that regulatory divergence can partly be recapitulated on the basis of DNA accessibility. This holds true, especially for *cis*-regulatory changes, where we found, that these are based on both, namely changes in DNA accessibility, as well as sequence divergence in orthologous regulatory regions. Comparing two different species pairs we confirm, that the amount of *cis*-regulatory divergence correlates with the phylogenetic distance in the *Drosophila melanogaster* subgroup. ASE expression analysis cannot reveal the causative genetic variants leading to differential expression, but with the combination of open chromatin datasets one can start to dissect the underlying genetic regulatory architecture. Our result that in general more *trans*-regulatory changes seem to underlie gene expression divergence between closely related species, calls for more tissue specific ASE studies in other animal

groups. It will be also interesting to reveal how the here described patterns deviate in other tissues, like the developing thoracic or leg imaginal discs to learn more about the context dependency of regulatory divergence.

5.5. Material and Methods

5.5.1. RNA-seq

The generation of RNA-seq datasets of developing eye-antennal discs (96h AEL) was performed as described for *D. melanogaster* in (Torres-Oliva et al., 2018). The same procedure was applied for datasets of *D. mauritiana* and *D. simulans*. In short, developing eye-antennal discs were dissected at 96h AEL for *D. melanogaster* and *D. mauritiana*. Please note, that in *D. simulans*, the morphogenetic furrow progressed a bit slower than in the other two species, therefore disc were dissected at 98h AEL, to ensure the same developmental time point (in the manuscript we still refer to '96h AEL' for the sake of clarity). To set up the hybrid crosses 400 *D. melanogaster* or *D. simulans* virgin females were crossed to 300 *D. mauritiana* males and the respective discs were dissected at 96h AEL. mRNA was extracted using the standard Trizol protocol and library preparation was prepared as described in (Torres-Oliva et al., 2018).

Differential expression analysis between parental strains and subsequent allele specific expression analysis (ASE) was performed by Dr. Torres-Oliva M. and is described in (Torres-Oliva, 2016).

We used the online tool Metascape (Zhou et al., 2019) to perform GO enrichment analysis for each group of genes (*cis*, *trans*, compensatory, conserved) in both pairwise settings (*D. melanogaster vs. D. mauritiana* and *D. simulans vs. D. mauritiana*.). In cases where more than 3000 genes were tested for enrichment, we randomly chose 3000 genes from the pool, since Metascape does not allow more genes as input.

5.5.2. ATAC-seq

5.5.2.1. ATAC-seq library preparation

For the generation of ATAC-seq datasets we followed (Buenrostro et al., 2013). Of all three species (*D. melanogaster, D. mauritiana* and *D. simulans*), developing eye-antennal discs were dissected in ice-cold PBS at 96h AEL. Please note, that in *D. simulans*, the morphogenetic furrow progressed a bit slower than in the other two species, therefore disc were dissected at 98h AEL, to ensure the same developmental time point (in the manuscript we still refer to '96h AEL' for

the sake of clarity). PBS was removed and exchanged for 50 µl lysis buffer (10 mM Tris-HCl (pH = 7.4); 10 mM NaCl; 3 mM MgCl2; 0.1 % IGEPAL). The mixture was pipetted several times up and down to lyse the cells and then split into micro centrifuge tubes. Centrifugation for 10 min at 500 g and 4 °C. The cell number was assessed in one of the samples and 50,000 to 80,000 nuclei were used in subsequent steps. The supernatant was removed and the pellet(s) dissolved in 47.5 µl 1X tagmentation buffer (20 mM Tris-CH3COOH (pH = 7.6); 10 mM MgCl2; 20 % (vol/vol) dimethylformamide) with 2.5 µl Tn5 Transposase and then incubated for 30 min at 37 °C. For purification we used the QIAGEN MinElute Kit and eluted in 10 µl Elution Buffer (10 mM Tris, pH = 8). For the PCR amplification was done as follows:

- 10 µl tagmented chromatin
- 10 µl H2O
- 2.5 µl Nextera PCR primer 1*
- 2.5 µl Nextera PCR primer 2**
- 25 µl NEBNext High-Fidelity 2X PCR Master Mix (Cat #M0541)

We used the following program:

(8) 72 °C 5 min
(9) 98 °C 30 sec
(10) 98 °C 10 sec
(11) 63 °C 30 sec
(12) 72 °C 1 min
(13) repeat 3-5 13 times
(14) hold at 4 °C

followed by another 2x purification step with the QIAGEN MinElute Kit: elution in 2 X 10 µl Elution Buffer (10 mM Tris, pH = 8).

* AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG

** Ad2.2_CGTACTAG CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
Ad2.3_AGGCAGAA CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4_TCCTGAGC CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5_GGACTCCT CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
Ad2.6_TAGGCATG CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
Ad2.7_CTCTCTAC CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT

5.5.2.2. Bioinformatics

We performed quality checks of the sequenced reads using FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The reads were trimmed, using Trimmomatic (version 0.36) (Bolger et al., 2014) appyling a sliding window trimming with the parameters `slidingwindow 4:15` and `minlen 30`. Trimmed reads were mapped to the *D. melanogaster* genome (version 6.13) after discarding the mitochondrial genome, using Bowtie2 (version 2.3.4.3) (Langmead et al., 2009), with the commands: `--no-unal` and `-X2000`. Samtools version 1.9 (Li et al., 2009) were subsequently used to convert the sam to bam files, and to sort and index bam files. We removed duplicates using PICARD (version 2.1.1, https://github.com/broadinstitute/picard) with default parameters and converted the resulted bam files to bed files. Reads were then centered as described in (Buenrostro et al., 2013). We used MACS2 (version 2.1.2) (Zhang et al., 2008, p. 2) with the following commands `-g dm --nomodel --shift -100 --extsize 200 -q 0.01 –bdg` to call significant peaks. We used the Integrated Genome Browser (IGB, (Freese et al., 2016)) to visualize the read depth and peaks. Peaks were annotated to the closest gene using the `annotatePeaks.pl` program from the HOMER software package (v4.8.3) using dm6 as genomic input.

5.5.3. Conversion of Coordinates

To compare the open chromatin landscape of all three sister species, we converted peak coordinates of *D. mauritiana* and *D. simulans* into *D. melanogaster* coordinates. This required to create custom liftOver files, also called chain files which are usually used to convert annotations from one genome version to the other (here from one species to another).

First, the *D. melanogaster* genome was indexed and each chromosome arm was saved separately as a .fasta file, using the `samtools faidx` command (Li et al., 2009). The same was done for the *D. mauritiana* strain-specific genome (TAM16). Each chromosome arm sequence was then split into chunks using the following command to ensure an efficient BLAT alignment and a `.lft` file was directly created using the size parameters: `faSplit –lift=Dmel_x.lft –oneFile size Dmel/dem-x.fasta 3000 dmel-x_chunks`, where `x` stands for each chromosome arm. The resulting sequence chunks were then aligned to the *D. mauritiana* genome sequences using the BLAT alignment tool (Kent, 2002) with default parameters and `.psl` files as output. The coordinates of the alignment were then changed to the *D. melanogaster* coordinate system using the `liftUp` tool (Hinrichs et al., 2006) and the `.lft` files created in the chromosome split step (see above). The converted alignments were

then chained together using the axtChain tool with the following parameters: `-psl –linearGap=medium –faQ –faT`. The resulting chain file (of each chromosome arm) were then combined and sorted using the `chainMergeSort` program. Using the chromosome sizes, co-called nets were created form chains using `chainNet` and subsequently `netChainSubset` was used to create over.chain files which are the files used for the coordinate conversion of peaks. This pipeline was adapted to both species, *D. mauritiana* and *D. simulans*. To convert peak coordinates, the liftOver tool was used with the parameter: `-minMatch=0.1`. The script can be found in the Appendix. Peaks which could not be converted were visualized using the Integrated Genome Browser (IGB) (Freese et al., 2016) and since nearly all mapped to the centromere regions of the respective chromosomes, we removed this regions from the *D. melanogaster* genome as well for further analysis.

5.5.4. Comparison of peak architectures

To get the consensus peak set of two sister species we used the `bedtools intersect` tool from the Bedtools toolset (version 2.24) (Quinlan, 2014; Quinlan and Hall, 2010) with the following parameters: `-wa -wb –wo`. Species specific peaks were extracted using the same tool with the `–v` parameter. All further analyses were carried out using R version 3.3.3 or 3.5.2 (R Development Core Team, 2008). Gene sets were combined according to the following criteria: Set1 – genes did not have a single overlapping peak between two species, Set2 – genes that had only overlapping peaks, and Set3 – genes that had overlapping and additional non-overlapping peaks. These genes were then overlapped with the information gained from our RNA-seq and ASE analysis, namely if the gene was differentially expressed between the two species, and if yes, which regulatory type was responsible for this differential expression. Fisher's exact tests was used to test for significances among the groups in the contingency tables.

5.5.5. Sequence alignments

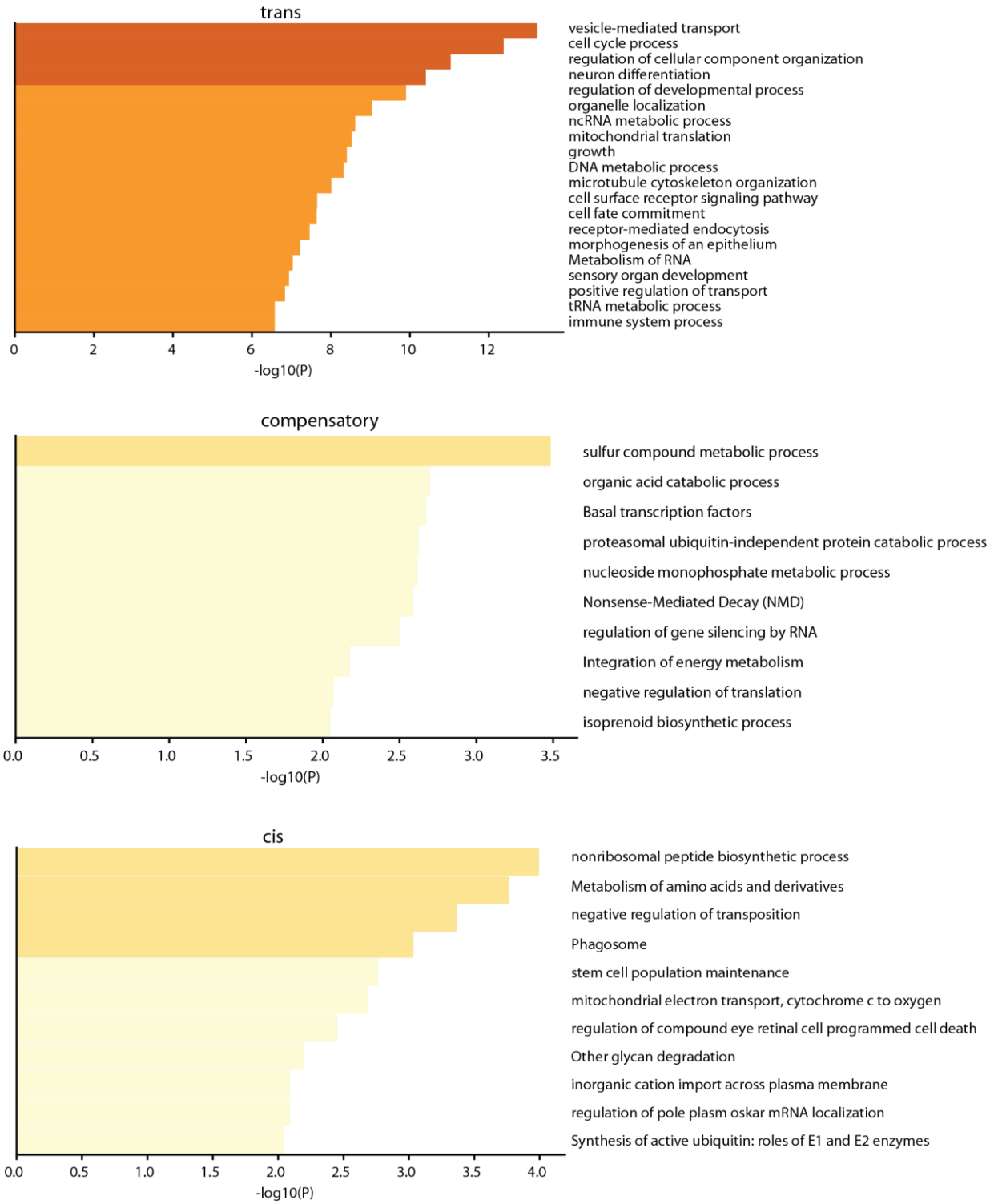
To get the sequences of homologous peaks, the Bedtools (version 2.24) `getfasta` programm was used with default parameters. Peak sequences of *D. mauritiana* were then used to build a BLAST database (Camacho et al., 2009) using the `-parse_seqids -dbtype nucl` parameters. blastn –db was then used to blast the peak sequences in the two comparisons, with parameters `-outfmt 6 -max_target_seqs 1 -evalue 0.01`. Wilcoxon Rank Sum test was applied to compare the percentage of identical matches between the groups (*cis*-,

trans-, conserved or compensatory genes). To test if intronic sequences and TSS/promoter sequences show differences in sequences conservation, we split all peak sequences in the groups according to the respective annotation and repeated the analysis separately for both groups.

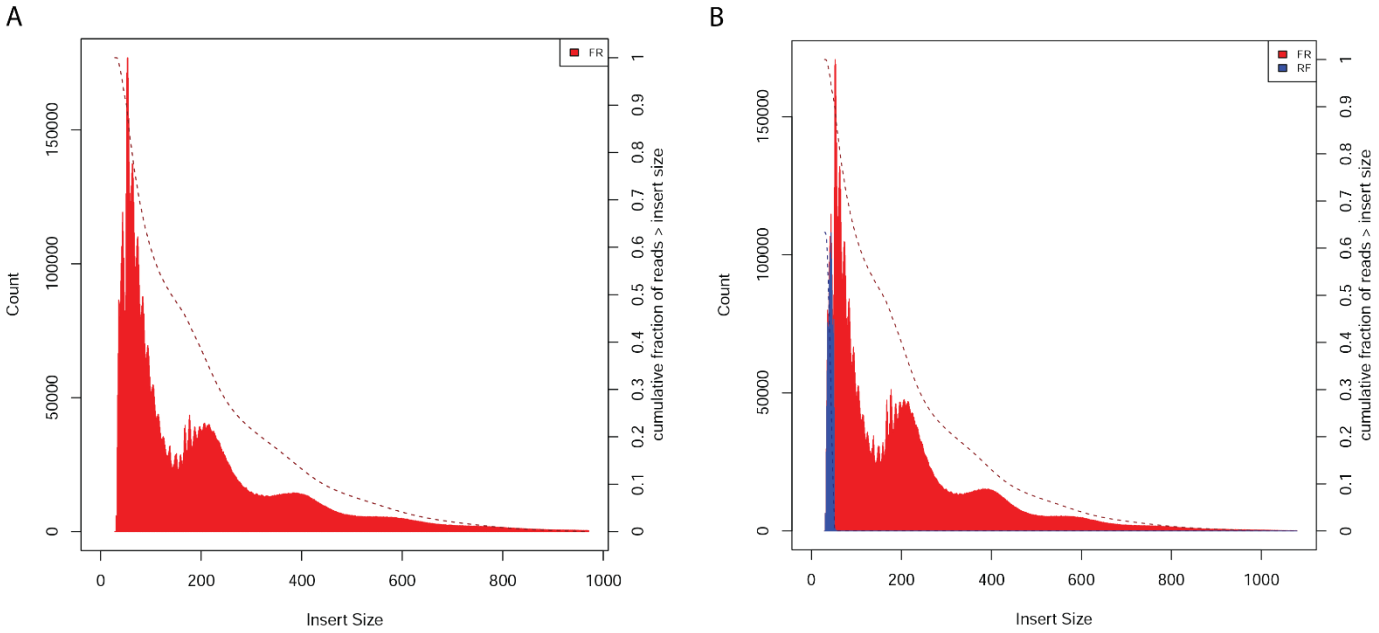
5.5.6. Overlap with DroID database

To test for regulatory divergence specifically in transcription factors, we downloaded all entries from the Transcription Factor TF - Gene Interaction data file in the DroId database (version 2015-2) (Yu et al., 2008), including 157462 interactions and 12323 genes. We then overlapped the transcription factors with our RNA-seq dataset to filter those that are expressed in eye-antennal disc. Of these we retrieved the then the information about potential differential expression of the transcription factors between species, by comparing them with our differential expression analysis and further checked if differential expression was due to *cis*- or *trans*-regulatory changes. We additionally checked if the transcription factors fall into Gene Set 1, 2, or 3, allowing us to categorize their regulatory landscape.

5.6. Supplementary Figures

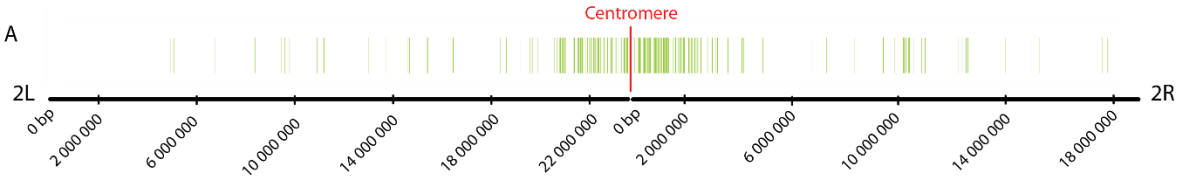


Supplementary Figure 21. GO enrichment analysis following ASE analysis between *D. simulans vs. D. mauritiana*. Genes that were differentially expressed due to *trans*-regulatory changes were enriched in morphogenetic, cell cycle, growth and developmental GO terms. Genes, showing compensatory regulation in the hybrids were enriched in more metabolic processes. Gene with *cis*-regulatory divergence showed enrichment in similar processes, namely biosynthetic and metabolic processes, but also in more eye-specific processes, like 'retinal cell programmed cell death'.



Supplementary Figure 22. Insert size distribution of ATAC-seq datasets of *D. mauritiana* and *D. simulans*. A. The insert size distribution of the *D. mauritiana* ATAC-seq dataset and **B.** of *D. simulans* show the same typical periodicity of ~200 bp as *the D. melanogaster* dataset (Figure 22B.).

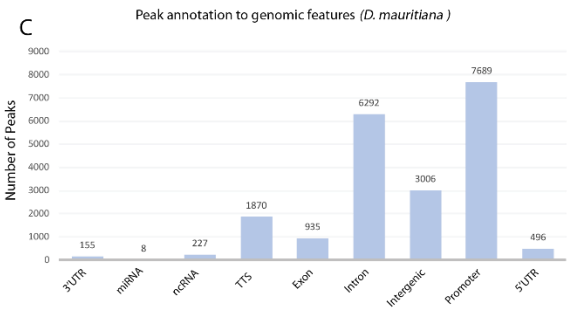
A



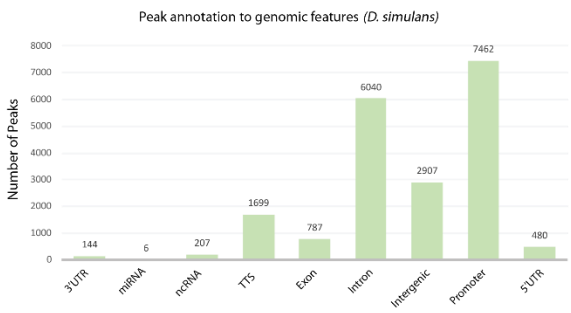
B

	chromosome	peaks excluded	region excluded
D.melanogaster	Chr-2L	0	0
	Chr-2R	305	< 4.100.000
	Chr-3L	203	> 24.000.000
	Chr-3R	183	< 4.160.000
	Chr-4	0	0
	Chr-X	734	> 18.270.000
D. mauritiana	Chr-2L	0	
	Chr-2R	4	
	Chr-3L	5	
	Chr-3R	13	
	Chr-4	0	
	Chr-X	0	
D. simulans	Chr-2L	0	
	Chr-2R	99	
	Chr-3L	133	
	Chr-3R	111	
	Chr-4	0	
	Chr-X	505	

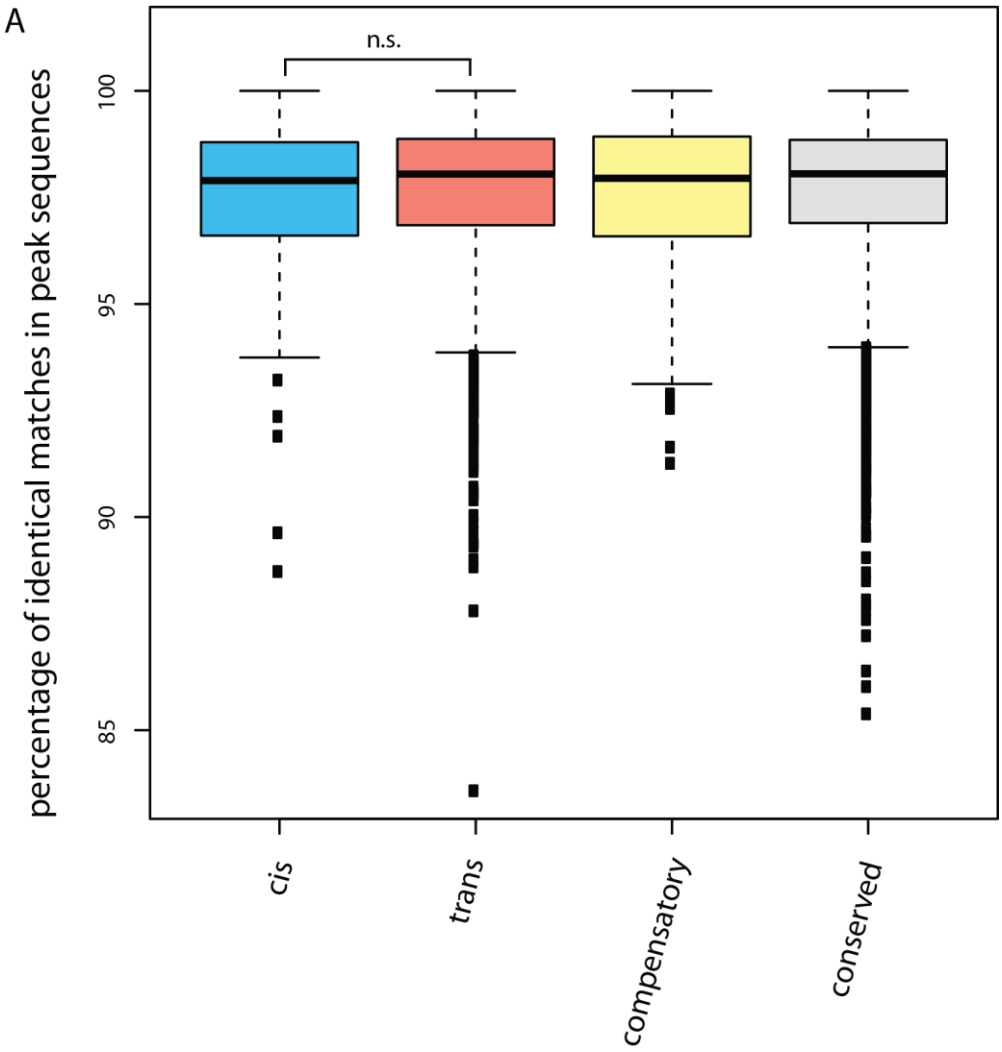
C



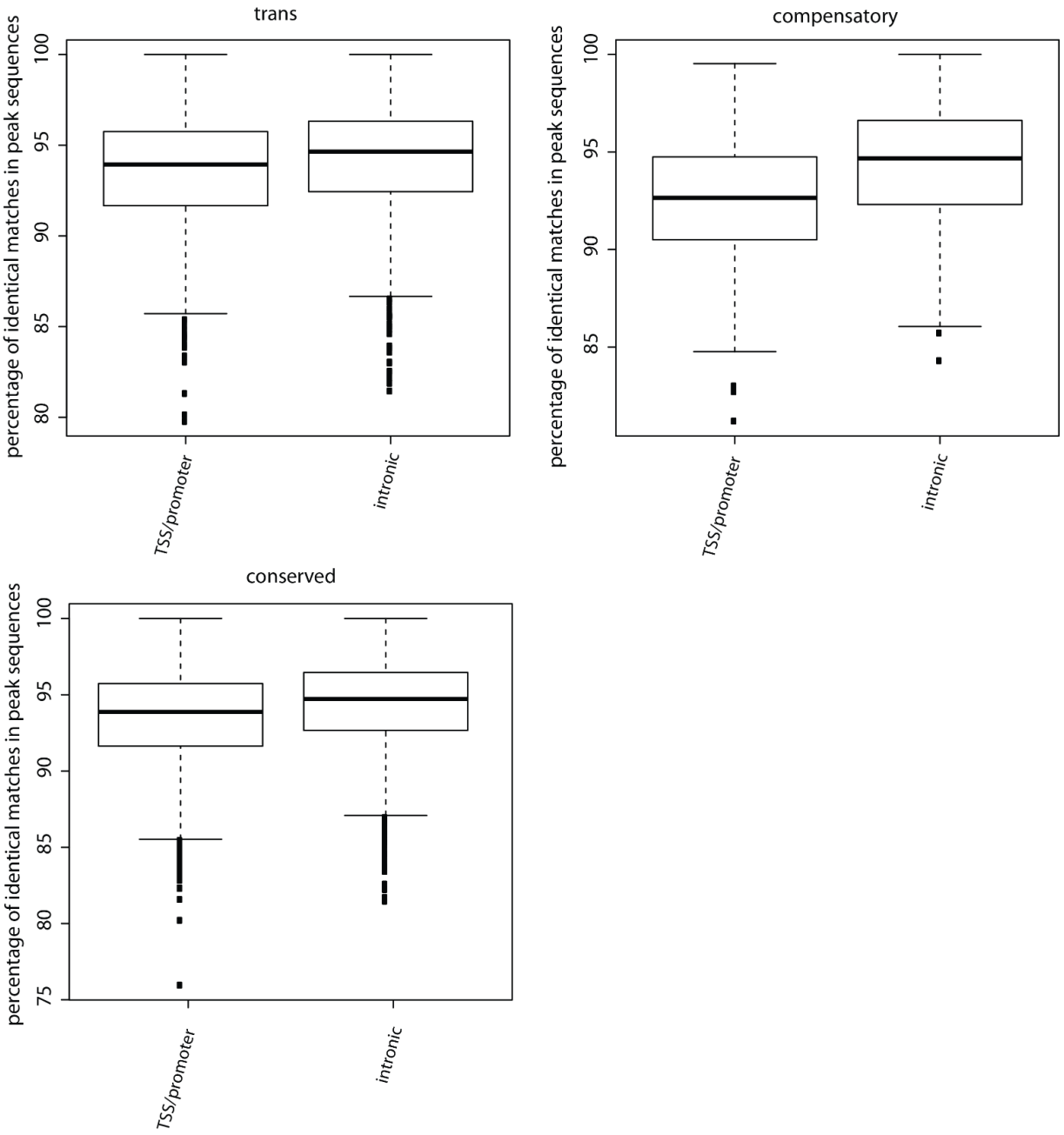
D



Supplementary Figure 23. Conversion of genomic coordinates. A. Peak coordinates that could not be converted from the *D.simulans* to the *D. melanogaster* genomic coordinate system mapped predominantly to the centromeric regions of the chromosomes. Shown here is the 2nd chromosome of *D. simulans*. **B.** Peaks that were excluded by filtering centromeric regions and peaks that did not map in each species. The last column in *D. melanogaster* lists the regions that were excluded for each chromosome arm (in bp). **C.** Converted peaks were annotated to gene features. The pattern is comparable to *D. melanogaster*, where also most peaks were annotated to promoter regions, followed by intronic regions and intergenic regions.



Supplementary Figure 24. A. The sequence divergence between peak sequences close to genes showing *cis*-regulatroy divergence or 'compensatory - genes' in the hybrid do not show significantly more sequence changes than conserved genes or genes with *trans*-regulatory divergence when *D. simulans* is compared to *D. mauritiana*.



Supplementary Figure 25. As shown for genes with *cis*-regulatory divergence between *D. melanogaster* and *D. mauritiana*, TSS\promoter peaks of genes from all divergence groups show a significantly higher sequence divergence than intronic peaks. Note that we provide all p-values between the pairwise comparisons in Supplementary Table 22.

5.7. Supplementary Tables

Supplementary Table 20. A. Fisher's Exact test (p-values) for pairwise comparisons of gene sets between *D. melanogaster vs. D. mauritiana* from Figure 25B. **B.** Fisher's Exact test (p-values) for pairwise comparisons of gene sets between *D. mauritiana vs. D. simulans* from Figure 25C.

A

Fisher's Exact test - Gene Sets (ad. Figure 4A)			
Set	**Comparison**		**p-value**
1	cis	trans	9.54E-03
1	cis	compensatory	3.83E-02
1	cis	conserved	1.57E-03
1	trans	compensatory	7.59E-01
1	trans	conserved	5.35E-01
1	compensatory	conserved	1.00E+00
2	cis	trans	8.12E-01
2	cis	compensatory	4.34E-01
2	cis	conserved	1.10E-02
2	trans	compensatory	2.28E-01
2	trans	conserved	4.76E-05
2	compensatory	conserved	1.31E-03
3	cis	trans	1.60E-07
3	cis	compensatory	4.92E-01
3	cis	conserved	4.12E-01
3	trans	compensatory	2.17E-04
3	trans	conserved	2.20E-16
3	compensatory	conserved	8.62E-01

B

Fisher's Exact test - Gene Sets (ad. Supplementary Figure 4A)			
Set	**Comparison**		**p-value**
1	cis	trans	5.16E-03
1	cis	compensatory	1.75E-02
1	cis	conserved	3.52E-04
1	trans	compensatory	7.05E-01
1	trans	conserved	4.45E-02
1	compensatory	conserved	7.73E-01
2	cis	trans	2.94E-02
2	cis	compensatory	1.53E-01
2	cis	conserved	9.16E-03
2	trans	compensatory	7.61E-01
2	trans	conserved	2.20E-16
2	compensatory	conserved	3.49E-05
3	cis	trans	3.75E-02
3	cis	compensatory	1.69E-01
3	cis	conserved	8.23E-01
3	trans	compensatory	7.24E-01
3	trans	conserved	2.63E-09
3	compensatory	conserved	7.45E-02

Supplementary Table 21. Wilcoxon Rank Sum Test (p-values) of sequence alignments between regulatory regions between *D. melanogaster* and *D. mauritiana* (Boxplots are shown in Figure 24A).

Wilcoxon Rank Sum Test - Sequence Alignment (ad. Figure 4B)		
Comparison		**p-value**
cis	trans	1.06E-12
cis	compensatory	2.40E-02
cis	conserved	1.07E-13
trans	compensatory	3.61E-04
trans	conserved	8.64E-01
compensatory	conserved	2.11E-04

Supplementary Table 22. Wilcoxon Rank Sum Test (p-values) of sequence alignments between intronic and TSS/promoter regulatory regions for each regulatory type. The boxplot for the *cis*-regulatroy changes is depicted in Figure 24B, whereas the three boxplots for *trans*-regulatry, compensatoy and conserved gene sets are shown in Supplementary Figure 25.

Wilcoxon Rank Sum Test -TSS/promoter vs. intronic (ad. Figure 4C and Supplementary Figure 5)	
regulatory divergence type	**p-value**
cis	9.19E-03
trans	7.81E-08
compensatory	1.48E-04
conserved	2.20E-16

Supplementary Table 23. Transcription Factors downloaded from the DroID database were first overlapped with the Gene Sets 1-3 (highly diverged regulatory regions, conserved regulatory regions and sligthly diverged regulatory regions) and second with the information about their divergence type. Most transcription factors show a highly conserved regulatory region and their expression levels are conserved as well between the species. If differentially expressed the genes are differentially expressed due to upstream *trans*-regulatory changes.

Transcription factors (DroID) in the gene sets					
Set 1		Set 2		Set 3	
Dmel_Dmau	Dsim_Dmau	Dmel_Dmau	Dsim_Dmau	Dmel_Dmau	Dsim_Dmau
ci	Jra	abd-A	abd-A	Abd-B	Antp
CTCF	sxc	Adf1	Abd-B	Antp	ap
E(spl)m5-HLH	tgo	Aef1	Adf1	ap	apt
gsb	twi	ara	Aef1	apt	bab1
hb		ato	ara	bab1	bin
Jra		bcd	ato	BEAF-32	br
sna		brm	bcd	bin	byn
tgo		Cf2	BEAF-32	br	chinmo
twi		Chro	brk	brk	cnc
		Cp190	brm	byn	ct
		CtBP	cad	**cad**	CtBP
trans		**D**	CBP	CBP	Dfd
compensatory		DCTN1-p150	Cf2	chinmo	Dll
cis		Dfd	Chro	cnc	Dsp1
conserved		disco	Cp190	CrebA	dsx
		dl	CrebA	ct	**E2f1**
		Doc1	D	Deaf1	EcR
		dpn	DCTN1-p150	Dll	Eip74EF
		Dref	Deaf1	dsx	en
		Dsp1	disco	E2f1	ftz-f1
		dwg	dl	EcR	grh
		dysf	Doc1	Eip74EF	gro
		E(spl)m8-HLH	dpn	en	gsb-n
		E(z)	dwg	ey	Hr39
		E2f2	**dysf**	ftz-f1	Hr46
		ems	E(spl)m5-HLH	grh	inv
		eve	E(spl)m8-HLH	gsb-n	lab
		exd	E(z)	Hr39	Mad
		fkh	E2f2	Hr46	Mef2
		gcm	ems	hth	oc
		gl	eve	inv	opa
		gro	exd	jumu	ovo
		h	ey	lab	Pdp1
		hkb	fkh	lz	ph-p
		Hsf	gcm	Mad	pnt
		ind	gl	**Mef2**	sbb
		insv	gsb	oc	shn
		kn	h	opa	srp
		kni	hb	ovo	Stat92E
		Kr	hkb	pan	sv
		Med	Hsf	Pax	toy
		mip120	hth	Pdp1	trx
		MTF-1	ind	ph-p	ttk
		Myb	insv	pnt	vnd
		NELF-B	jumu	sbb	z
		nub	kn	shn	zfh1
		Pc	kni	Stat92E	
		pho	Kr	sv	
		phol	lz	toy	
		pnr	Med	ttk	
		prd	mip120	z	
		sd	MTF-1		
		sens	Myb		
		Sfmbt	NELF-B		
		slbo	nub		
		slp1	pan		
		Snr1	Pax		
		so	Pc		
		srp	pho		
		Su(H)	phol		
		su(Hw)	pnr		
		TfIIB	prd		
		tin	sd		
		tll	sens		
		Top2	Sfmbt		
		Trl	slbo		
		trx	slp1		
		tup	Snr1		
		Ubx	so		
		vnd	Su(H)		
		vvl	su(Hw)		
		zfh1	TfIIB		
			tin		
			tll		
			Top2		
			Trl		
			tup		
			Ubx		
			vvl		

5.8. Appendix

```
###The script was written to generate chain files to convert open-chromatin
peak coordinates from D. mauritiana to D. melanogaster
###The same pipeline was applied to D. simulans

#!/bin/bash
#script was adapted from http://blog.windhager.io/2016/10/21/creating-
liftover-chain-files/

# requires UCSC genome browser 'kent' bioinformatic utilities

module load EMBOSS/6.5.7 UCSC/20160601

mkdir psl
mkdir chain
mkdir net

#get all the chromosomes in extra files

mkdir Dmel
mkdir Dmau

samtools faidx dmel-all-chromosome-r6.13_woMito.fasta 2L > Dmel/dmel-
2L.fasta
samtools faidx dmel-all-chromosome-r6.13_woMito.fasta 2R > Dmel/dmel-
2R.fasta
samtools faidx dmel-all-chromosome-r6.13_woMito.fasta 3L > Dmel/dmel-
3L.fasta
samtools faidx dmel-all-chromosome-r6.13_woMito.fasta 3R > Dmel/dmel-
3R.fasta
samtools faidx dmel-all-chromosome-r6.13_woMito.fasta 4 > Dmel/dmel-4.fasta
samtools faidx dmel-all-chromosome-r6.13_woMito.fasta X > Dmel/dmel-X.fasta

samtools faidx TAM16_strainspecificGenome_woMito.fasta Dmau_2L > Dmau/dmau-
2L.fasta
samtools faidx TAM16_strainspecificGenome_woMito.fasta Dmau_2R > Dmau/dmau-
2R.fasta
samtools faidx TAM16_strainspecificGenome_woMito.fasta Dmau_3L > Dmau/dmau-
3L.fasta
samtools faidx TAM16_strainspecificGenome_woMito.fasta Dmau_3R > Dmau/dmau-
3R.fasta
samtools faidx TAM16_strainspecificGenome_woMito.fasta Dmau_4 > Dmau/dmau-
4.fasta
samtools faidx TAM16_strainspecificGenome_woMito.fasta Dmau_X > Dmau/dmau-
X.fasta

# split new sequences for efficient BLAT alignment

faSplit -lift=Dmel_2L.lft -oneFile size Dmel/dmel-2L.fasta 3000 Dmel/dmel-
2L_chunks
faSplit -lift=Dmel_2R.lft -oneFile size Dmel/dmel-2R.fasta 3000 Dmel/dmel-
2R_chunks
faSplit -lift=Dmel_3L.lft -oneFile size Dmel/dmel-3L.fasta 3000 Dmel/dmel-
3L_chunks
faSplit -lift=Dmel_3R.lft -oneFile size Dmel/dmel-3R.fasta 3000 Dmel/dmel-
3R_chunks
faSplit -lift=Dmel_4.lft -oneFile size Dmel/dmel-4.fasta 3000 Dmel/dmel-
4_chunks
faSplit -lift=Dmel_X.lft -oneFile size Dmel/dmel-X.fasta 3000 Dmel/dmel-
X_chunks
```

```
# align resulting sequence chunks to old sequence, which is in my case the
Dmau genome

/home/uni05/ebuchbe/Programme/./blat
Dmau/TAM16_strainspecificGenome_woMito.fasta Dmel/dmel-2L_chunks.fa
psl/chr2L_blat_param.psl &
/home/uni05/ebuchbe/Programme/./blat
Dmau/TAM16_strainspecificGenome_woMito.fasta Dmel/dmel-2R_chunks.fa
psl/chr2R_blat_param.psl &
/home/uni05/ebuchbe/Programme/./blat
Dmau/TAM16_strainspecificGenome_woMito.fasta Dmel/dmel-3L_chunks.fa
psl/chr3L_blat_param.psl &
/home/uni05/ebuchbe/Programme/./blat
Dmau/TAM16_strainspecificGenome_woMito.fasta Dmel/dmel-3R_chunks.fa
psl/chr3R_blat_param.psl &
/home/uni05/ebuchbe/Programme/./blat
Dmau/TAM16_strainspecificGenome_woMito.fasta Dmel/dmel-4_chunks.fa
psl/chr4_blat_param.psl &
/home/uni05/ebuchbe/Programme/./blat
Dmau/TAM16_strainspecificGenome_woMito.fasta Dmel/dmel-X_chunks.fa
psl/chrX_blat_param.psl &

# change alignment coordinates to parent coordinate system according to LFT
file
#LiftUp can convert coordinates in most annotation files. It can add to
positions and change the chromosome part of those files. It's main input is
the lift-file that specifies how to convert the coordinates.
liftUp -pslQ psl/chr2L.psl Dmel_2L.lft warn psl/chr2L_blat_param.psl
liftUp -pslQ psl/chr2R.psl Dmel_2R.lft warn psl/chr2R_blat_param.psl
liftUp -pslQ psl/chr3L.psl Dmel_3L.lft warn psl/chr3L_blat_param.psl
liftUp -pslQ psl/chr3R.psl Dmel_3R.lft warn psl/chr3R_blat_param.psl
liftUp -pslQ psl/chr4.psl Dmel_4.lft warn psl/chr4_blat_param.psl
liftUp -pslQ psl/chrX.psl Dmel_X.lft warn psl/chrX_blat_param.psl

# chain together alignments from PSL files

axtChain -psl -linearGap=medium -faQ -faT psl/chr2L.psl
Dmau/TAM16_strainspecificGenome_woMito.fasta Dmel/dmel-2L.fasta
chain/chr2L_axtChain.chain
axtChain -psl -linearGap=medium -faQ -faT psl/chr2R.psl
Dmau/TAM16_strainspecificGenome_woMito.fasta Dmel/dmel-2R.fasta
chain/chr2R_axtChain.chain
axtChain -psl -linearGap=medium -faQ -faT psl/chr3L.psl
Dmau/TAM16_strainspecificGenome_woMito.fasta Dmel/dmel-3L.fasta
chain/chr3L_axtChain.chain
axtChain -psl -linearGap=medium -faQ -faT psl/chr3R.psl
Dmau/TAM16_strainspecificGenome_woMito.fasta Dmel/dmel-3R.fasta
chain/chr3R_axtChain.chain
axtChain -psl -linearGap=medium -faQ -faT psl/chr4.psl
Dmau/TAM16_strainspecificGenome_woMito.fasta Dmel/dmel-4.fasta
chain/chr4_axtChain.chain
axtChain -psl -linearGap=medium -faQ -faT psl/chrX.psl
Dmau/TAM16_strainspecificGenome_woMito.fasta Dmel/dmel-X.fasta
chain/chrX_axtChain.chain

# combine and sort chain files
#chainSort chain/chr1_axtChain.chain chain/chr2_axtChain.chain
chain/chr3_axtChain.chain | chainSplit chain stdin
chainMergeSort chain/chr2L_axtChain.chain chain/chr2R_axtChain.chain
chain/chr3L_axtChain.chain chain/chr3R_axtChain.chain
chain/chr4_axtChain.chain chain/chrX_axtChain.chain| chainSplit chain stdin
```

```
# determine chromosome sizes
faToTwoBit Dmau/dmau-2L.fasta Dmau/dmau-2L.2bit
faToTwoBit Dmau/dmau-2R.fasta Dmau/dmau-2R.2bit
faToTwoBit Dmau/dmau-3L.fasta Dmau/dmau-3L.2bit
faToTwoBit Dmau/dmau-3R.fasta Dmau/dmau-3R.2bit
faToTwoBit Dmau/dmau-4.fasta Dmau/dmau-4.2bit
faToTwoBit Dmau/dmau-X.fasta Dmau/dmau-X.2bit

faToTwoBit Dmel/dmel-2L.fasta Dmel/dmel-2L.2bit
faToTwoBit Dmel/dmel-2R.fasta Dmel/dmel-2R.2bit
faToTwoBit Dmel/dmel-3L.fasta Dmel/dmel-3L.2bit
faToTwoBit Dmel/dmel-3R.fasta Dmel/dmel-3R.2bit
faToTwoBit Dmel/dmel-4.fasta Dmel/dmel-4.2bit
faToTwoBit Dmel/dmel-X.fasta Dmel/dmel-X.2bit

{ twoBitInfo Dmau/dmau-2L.2bit stdout; twoBitInfo Dmau/dmau-2R.2bit stdout;
twoBitInfo Dmau/dmau-3L.2bit stdout; twoBitInfo Dmau/dmau-3R.2bit stdout;
twoBitInfo Dmau/dmau-4.2bit stdout; twoBitInfo Dmau/dmau-X.2bit stdout; } >
Dmau/chrom.sizes
{ twoBitInfo Dmel/dmel-2L.2bit stdout; twoBitInfo Dmel/dmel-2R.2bit stdout;
twoBitInfo Dmel/dmel-3L.2bit stdout; twoBitInfo Dmel/dmel-3R.2bit stdout;
twoBitInfo Dmel/dmel-4.2bit stdout; twoBitInfo Dmel/dmel-X.2bit stdout; } >
Dmel/chrom.sizes

# make alignment nets out of chains
mkdir net
chainNet chain/Dmau_2L.chain Dmau/chrom.sizes Dmel/chrom.sizes
net/chr_2L.net /dev/null
chainNet chain/Dmau_2R.chain Dmau/chrom.sizes Dmel/chrom.sizes
net/chr_2R.net /dev/null
chainNet chain/Dmau_3L.chain Dmau/chrom.sizes Dmel/chrom.sizes
net/chr_3L.net /dev/null
chainNet chain/Dmau_3R.chain Dmau/chrom.sizes Dmel/chrom.sizes
net/chr_3R.net /dev/null
chainNet chain/Dmau_4.chain Dmau/chrom.sizes Dmel/chrom.sizes net/chr_4.net
/dev/null
chainNet chain/Dmau_X.chain Dmau/chrom.sizes Dmel/chrom.sizes net/chr_X.net
/dev/null

# create over.chain
netChainSubset net/chr_2L.net chain/Dmau_2L.chain
chain/Dmau_2L_subset.chain
netChainSubset net/chr_2R.net chain/Dmau_2R.chain
chain/Dmau_2R_subset.chain
netChainSubset net/chr_3L.net chain/Dmau_3L.chain
chain/Dmau_3L_subset.chain
netChainSubset net/chr_3R.net chain/Dmau_3R.chain
chain/Dmau_3R_subset.chain
netChainSubset net/chr_4.net chain/Dmau_4.chain chain/Dmau_4_subset.chain
netChainSubset net/chr_X.net chain/Dmau_X.chain chain/Dmau_X_subset.chain
cat chain/Dmau_2L_subset.chain chain/Dmau_2R_subset.chain
chain/Dmau_3L_subset.chain chain/Dmau_3R_subset.chain
chain/Dmau_4_subset.chain chain/Dmau_X_subset.chain > over_DmauToDmel.chain

rm -rf psl chain net

# do the coordinate conversion with liftOver
# Usage:
# liftOver oldFile map.chain newFile unMapped

##liftOver the already called peaks for Dmau for comparison
```

```
module load EMBOSS/6.5.7 UCSC/20160601

liftOver -minMatch=0.1 TAM_96hA_peaks.bed over_DmauToDmel.chain
TAM_96hA_peaks_mapped.bed TAM_96hA_peaks_unmapped.bed &

###grep the unmapped peaks in Dmau for visualization

grep "Dmau" TAM_96hA_peaks_unmapped.bed >
TAM_96hA_peaks_unmapped_IGBinput.bed
```

6. General Discussion and Outlook

Evolutionary changes in phenotypes, including adult morphologies, life history or physiological traits are a prerequisite for a constant adaptation to an ever-changing environment and the result of heritable mutations in the genome. For variation in adult morphological structures it is widely accepted that such mutations often affect the developmental programs underlying the formation of the respective structures. Building a complex organism requires that an initially single cell differentiates into various cell types that make up a variety of tissues and eventually form functional organs. The instructions for these developmental processes are encoded in the genome and translated through stage- and tissue-specific gene expression, that allows a cell or a group of cells to acquire a specific fate. Consequently, a major goal of biological studies is to understand how a given genotype translates - on a molecular level - into relevant phenotypes ('genotype to phenotype map'). For morphological traits, the application of comparative developmental approaches has been proven to be a powerful way to achieve this goal.

6.1. Integration of different datasets in comparative biological studies

Historically, the relationships between animal lineages were often reconstructed by the comparison of adult morphological features (e.g. (Snodgrass, 1938)), resulting in numerous descriptions of morphological phenotypes and traits in a variety of organisms. Advances in molecular techniques and the establishment of genetic tools allowed a shift from comparative and descriptive studies, towards a more experimental discipline that made it possible to verify phylogenetic relationships on a molecular level. However, only the advent of high throughput sequencing technologies revolutionized the way to reconstruct such phylogenies (e.g. (Dunn et al., 2008; Oakley et al., 2013) for a very recent study see: (Laumer et al., 2019)). Approaches, that combine morphological data and genomic approaches have been used to resolve for example the relationships of Squamata (comprising lizards, snakes and amphisbaenia), including fossil taxa (Reeder et al., 2015) or to address the evolution of larger groups, like all deuterostomes (Swalla and Smith, 2008). Whole genome sequencing and/or transcriptomics also have the power to reveal major ancient evolutionary events like whole genome duplications, allowing for instance subsequent comparison of gene content and syntenies (eg. (Dehal and Boore, 2005; Schwager et al., 2017; Singh et al., 2015)). The finding that spiders and scorpions share an ancient genome duplication supports their close relationship compared to

arachnids that do not show signatures of this duplication (Schwager et al., 2017). Apart from gaining more insights into phylogenetic relationships, such data are also highly valuable in answering questions about phenotypic evolution, including neofunctionalization of genes and the emergence of evolutionary novelties (e.g. (Moriyama and Koshiba-Takeuchi, 2018; Turetzek et al., 2017, 2016)).

Apart from the comparison of adult morphologies, classical comparative developmental approaches like the analysis of, for instance, Hox genes in several lineages, have brought major insights into the evolution of body plans (e.g. (Akam, 1995; Akam et al., 1994; Garcia-Fernàndez and Holland, 1994)). Tarazona and colleagues recently used the cuttlefish *Sepia officinalis* to study if the developmental processes underlying appendage development are conserved among Bilaterians. They could indeed find that, despite legs of vertebrates, arthropods and cephalopods not being homologous structures, the 'developmental mechanisms' of appendage formation seem to be highly conserved (Prpic, 2019; Tarazona et al., 2019).

Comparative embryology resulted in the suggestion that vertebrate embryogenesis goes through highly conserved stages, so-called phylotypic stages. Haeckel proposed, based on his observations that species look exceptionally similar during certain stages of embryonic development, his 'biogenetic law', suggesting that the phylogeny is recapitulated during development of an organism (Haeckl, 1879, 1867; Losos, 2014). Even though it is clear nowadays that the biogenetic law does not reflect reality, gene expression data has indeed shown that the transcriptome expressed at defined stages of zebrafish or *Drosophila* development shows signatures of an hourglass (Domazet-Lošo and Tautz, 2010). While molecular tools were for a long time only available for a few model systems, affordable sequencing technologies facilitated in recent years the establishment of genomic resources not only for classical, but also emerging model systems (Ellegren, 2014). Sequencing the genome and analyzing open chromatin datasets of *Branchiostoma lanceolatum*, the Mediterranean amphioxus, recently revealed that gene expression and the *cis*-regulatory architecture are highly conserved in all chordates during certain stages of development but showed that this phylotypic stage (i.e. the time point showing minimal transcriptomics divergence) between *Branchiostoma* and other vertebrates occurs at a slightly earlier time point compared to vertebrates (Marlétaz et al., 2018). Overall, it is relatively easy these days to provide a detailed description of the genotype for many different organisms for which detailed anatomical data

has been revealed over the years. But even though many genomes are sequenced, and the morphology and development of many organisms are described, it remains often elusive to recapitulate how the genomic information is used to define the adult phenotype. Further, mechanistic insights are in most cases missing. In summary, one can assume that the integration of morphological, developmental and molecular datasets allows comprehensive insights in phylogenetic relationships (Lee and Palci, 2015) and the genotype-phenotype map.

Here I argue that the combination of various detailed datasets provides the means to establish genotype-phenotype associations. First, a thorough understanding of the phenotype of interest is necessary. Additionally, for morphological traits it is highly informative to gain insights into developmental differences. Second, a comprehensive overview of the gene content and the genome size/organization is helpful. This can be achieved by generating transcriptome and genome datasets. Third, a correlation between the genotype and the phenotype must be established. If closely related species, that do not yet result in sterile offspring, are studied this can be done by quantitative genetics approaches such as QTL mapping or GWAS. Also, gene expression has been extensively used as an intermediate phenotype, backed up by the fact, that many mapped variances were described that influence gene expression (e.g. (Chan et al., 2010; Coyle et al., 2007; Cresko et al., 2004; Dixon et al., 2007; Gilad et al., 2008; Jia and Xu, 2007; Rockman and Kruglyak, 2006; Shapiro et al., 2004)). Hence, establishing gene expression differences between species (independent of the phylogenetic distance) allows identifying candidate genes responsible for morphological diversification. For morphological traits, such approaches are most powerful if they are combined with developmental data and if they are studied throughout development.

In each chapter of this work we used a combination of different datasets to connect phenotypes on several levels. In Chapter I, we applied microscopy techniques like scanning electron microscopy (SEM) and transmission electron microscopy (TEM) to analyze the function of pleuropodia in *Schistocerca gregaria* and connected the obtained insights with stage specific gene expression datasets. This allowed us to study long standing questions about the function of these organs, revealing potential new functions and as discussed below, holds the possibility to ask more general questions about developmental processes. In Chapter II and III, we used the model species *Drosophila melanogaster* and its sister species *Drosophila mauritiana* and *Drosophila simulans* to understand how complex traits like organ size and shape can evolve. As

in Chapter I, we used a comparative transcriptomic dataset as an intermediate phenotype to link this genetic readout to observable morphological changes. We applied geometric morphometrics to quantitatively compare adult head shapes and developed a semi-automated method to count individual ommatidia of single compound eyes. By adding a comparative ATAC-seq dataset, representing stage and tissue specific open chromatin landscapes, we were able to gain more genome-wide insights into the evolution of gene expression divergence and subsequently eye size and head shapes in these three closely related species.

The different types of transcriptomics and functional genomics datasets that I generated, will allow in the future to gain insights on a more global GRN level, going beyond a gene-centric approach. In the following two sections I will argue that a GRN-centric view will further result in new insights into development and phenotypic evolution.

6.2. Comparative gene expression studies and gene regulatory networks in development

Research in Evo-Devo has established that the development of diverse organisms as well as organs and tissues is based on a limited set of developmental genes, so-called 'toolkit genes' (Carroll et al., 2001). Intriguingly, many of these factors are highly conserved in different lineages (e.g. (Halder et al., 1995; King and Wilson, 1975)). Therefore, a central question is how this limited set of genes can control the development of different cell types and tissues? It is widely accepted nowadays, that differential expression of these genes and rewiring of regulatory interactions underlies the generation of differences between cells types and subsequently organs and that the proper development of organs and structures relies heavily on the correct temporal and spatial expression of genes. One of the best described examples exemplifying this is the development of the *Drosophila* nervous system. Initially identical precursor cells start to express distinct transcription factors in a spatially and temporally defined manner, leading to the formation of different neural identities (e.g. (Homem and Knoblich, 2012; Karcavich, 2005; Technau et al., 2006)). A great model to study how gene expression distinguishes organs are serially homologous structures, such as the insect appendages. We studied pleuropodia in the locust *S. gregaria*, small glandular structures that are apparent at the first abdominal segment of many insect embryos and are thought to be serially homologous to embryonic leg buds (Bennett et al., 1999; Lewis et al., 2000; Machida, 1981). These transient organs eventually mature and gain specific functions during

embryogenesis, but in contrast to other appendages they degenerate already before hatching of the embryo (Bullière, 1970; Louvet, 1975, 1973; Stay, 1977) . The comparison of developing leg buds and pleuropodia is therefore a valuable model to address the question of how differences in development, morphology and function of initially similar structures can arise. In our study we showed that especially in the early stages of embryonic development, legs and pleuropodia are not only morphologically extremely similar, but that this similarity is also recapitulated on a transcriptomic level. Genes which are known to be involved in leg development, for instance *distal-less* (*dll*) are also active in pleuropodia (Lewis et al., 2000; Yamamoto et al., 2004). In later stages we found that gene expression patterns become more and more divergent and gene set enrichment showed that the functional annotation of expressed transcripts gets more and more tissue specific. Our combinatorial approach revealed that pleuropodia of *S. gregaria* are indeed directly involved in the breakdown of the serosal cuticle and subsequently in the hatching of the insect embryo, supporting the result drawn by Slifer already 1937 (H. Slifer, 1937). Surprisingly, our GO-term enrichment analysis of differentially expressed genes points towards a role of pleuropodia in insect embryonic immunity, a function that is usually conferred by the extraembryonic serosa (Jacobs et al., 2014). It remains to be shown functionally if the pleuropodia take over immune protection of the embryo after degeneration of the serosa upon dorsal closure (Konopová et al., 2019; Panfilio, 2008). Overall, we demonstrate that the combination of thorough phenotyping of developing structures with the analysis of differential expression levels as an intermediate phenotype allows to gain major insights into function and developmental processes of embryonic structures.

While gene expression catalogs of various organs and expression dynamics of individual genes are being established for more and more developmental processes, it is not yet completely resolved how developmental genes are regulated in a tissue and stage specific manner. In recent years it became clear that the regulation of genes is not a simple hierarchical process but rather defined by an intricate interplay of gene products. These interactions are usually represented as so-called gene regulatory networks (GRNs) which describe genes or their products (transcription factors and other proteins) as nodes and the interaction among these (i.e. genetic interactions) as edges (Davidson and Levine, 2008; Thompson et al., 2015). GRNs provide therefore a logic and comprehensible cascade of the underlying developmental program. RNA profiling has been proposed as one of the main experimental procedures in

reconstructing GRNs, since it provides the possibility to collect all nodes that theoretically have to be considered in the respective GRN (shown for instance in (Sonawane et al., 2017), reviewed in (Thompson et al., 2015)). This allows to mathematically describe global properties of biological networks: One hallmark seems to be, that GRNs are so-called scale-free networks, meaning that the majority of the nodes is poorly connected and that we can only find a few highly connected nodes, also called hubs (Barabasi and Albert, 1999). Other measures, like node betweenness can further provide information about the role of single nodes in the global network architecture (e.g.(Koschützki and Schreiber, 2008)). One question that arises is how GRNs confer tissue specificity, i.e. when and where initially similar GRNs change and get rewired. By comparing tissue specific GRNs from adult human organs, Sonawane and colleagues found a rather low number of tissue-specific transcription factors (Sonawane et al., 2017). They show that functional specificity is primarily ensured by tissue specific interactions and that the expression of transcription factors is less well correlated with the regulation of functions in specific organs. Instead, tissue specific target gene expression is rather accomplished by context dependent paths throughout the network (Sonawane et al., 2017). Also, during eye-and head development in *Drosophila*, it was shown, that the same genes are able to exert different functions, mainly via rewiring of existing nodes (Palliyil et al., 2018). Palliyil and colleagues suggested that the retinal determination gene network is first important for overall growth of the complete eye-antennal disc, whereas later on, it specifically promotes retinal development. The fact that GRNs are constantly rewired during development provides an explanation, how morphological diversification can be achieved despite the developmental toolkit genes being not only expressed in one organ but are crucial for the proper development of several structures.

To dissect in more detail how and when such a rewiring takes place, our comparative dataset of *Schistocerca* pleuropodia and legs provides an excellent starting point. The already existing transcriptomic dataset allows to deduce which nodes are present and will need to be considered for the reconstruction of the respective tissue and stage specific GRN. The generation of a GRN depends though not only the information which genes have to be considered as nodes, but one also has to establish where to draw the edges. To add the edges globally it will be necessary to combine the RNA-seq dataset with, for instance, open chromatin datasets. ATAC-seq allows to search for transcription factor binding motifs in accessible and therefore potential *cis*-regulatory regions in the whole genome. This can be used to predict

direct genetic interactions between transcription factors and their target genes. While such a dataset remains to be established for pleuropodia, preliminary results of our ATAC-seq dataset of the eye-antennal disc suggests that the open chromatin landscape is dependent on the temporal context, since we found a number of stage specific peaks at each of the studied time points (72h, 96h and 120 AEL, data not shown). This is consistent with other studies conducted in *Drosophila*, that showed that the opening of regulatory regions is highly dynamic between stages during embryogenesis and also during larval stages (McKay and Lieb, 2013). Interestingly, the same study revealed also, that the accessibility of regulatory sequences in different developing appendages of *Drosophila* is exceptionally similar. The small number of tissue-specific open chromatin regions were annotated as regulatory regions of tissue specific master regulators (McKay and Lieb, 2013). In contrast, other studies found a highly unique and cell specific open chromatin landscape, for example in rods of murine retinas (Hughes et al., 2017). Hence, it will be interesting to investigate how changes in gene expression dynamics correlate with the open chromatin landscape in developing pleuropodia and legs. By comparing the output of this analysis between legs and pleuropodia at different stages, one might eventually be able to pinpoint the tissue and stage specific rewiring of the GRN that underlies the morphological differentiation of initially similar structures into two distinct organs. Overall, the combination of transcriptomic datasets with open chromatin datasets allows to tackle the question, how the rewiring of GRNs might be realized – on the level of chromatin accessibility and gene expression.

How the rewiring of existing GRNs is affecting direct gene interactions and impacts gene regulation on a mechanistic level, requires focusing on distinct nodes and edges. Apart from revealing global properties of biological networks, it has been shown, that GRNs are composed of smaller interaction entities or so-called circuits, which ensure certain gene expression outputs, like robustness or stochasticity (reviewed in (MacNeil and Walhout, 2011)). These are interactions between only a few nodes describe for instance feed-forward loops, autoregulatory loops or feed-back loops (reviewed in (MacNeil and Walhout, 2011)). In Chapter II of this work, we showed for the first time on a transcriptomics and protein level, that the co-factor of Pnr, called Ush is expressed in the eye-antennal disc. Subsequent validation of protein location and perturbation of *pnr* and *ush* expression levels allowed us in the following to analyze the small regulatory module of these two factors in more detail. Our results hint towards an auto-regulatory loop of Pnr, which is most probably kept in balance via the repressing function

of the heterodimer Pnr/Ush. Additionally, an activating role of Pnr on the expression of *ush* is very likely, and together with the repressing function of Pnr/Ush on *pnr* expression this interaction might represent a feed-back loop. These interactions are highly similar to the ones described for the developing wing imaginal disc (Fromental-Ramain et al., 2010). In their earlier work, Fromental-Ramain and colleagues additionally showed, that two isoforms of Pnr (Pnr-A and Pnr-B) are differentially expressed (Fromental-Ramain et al., 2008). We could confirm on the basis of quantitative real-time PCR and RNA-seq that, similar to the wing disc, *pnr-A* is not or only weakly expressed in the developing eye-antennal disc (data not shown). Thus, it remains to be analyzed, if also in these imaginal discs, the isoforms take over a distinct function, which would eventually refine the understanding of this regulatory module. Overall, the combination of RNA-seq with classical genetic tools can be used to define these small circuits which provide further information about direct transcriptional interactions on a more mechanistic level.

In summary, implementing a GRN centric view in developmental studies has great potential to broaden our current understanding of the molecular control of developmental processes. Especially the analysis of stage- and tissue-specific regulatory modules allows to understand mechanistically how a limited number of developmental gene products governs the formation of different tissues and organs.

6.3. Evolution of gene regulatory networks

Up to now we established, that the development of distinct organ fates relies on differential wiring of GRNs and consequently on tissue and stage specific gene expression. Since the GRN architecture and the transcriptomic landscape is highly variable across different serially homologous organs, it is as well conceivable that such variation also underlies the evolution of adult organs. Assuming that changes in developmental GRNs cause variation in adult structures requires therefore to study how such GRNs can change and evolve, primarily via the loss of existing edges or the connection of previously unconnected nodes. In all cases, the readout of such changes is reflected in changes of the transcriptional landscape (Thompson et al., 2015).

In our work we assumed that the GRNs that underlie the development of head structures in *D. melanogaster* are highly conserved in its closely related sister species, *D. simulans* and *D. mauritiana*. Nevertheless, *D. melanogaster* and *D. mauritiana* differ extensively in their eye size and head shapes, and even though the genetic architecture of such complex

traits has been started to be revealed (Arif et al., 2013a; Gaspar et al., 2019; Norry and Gomez, 2017), we lack - in most cases (for an exception see: (Ramaekers et al., 2018)) - functional data that validate single candidate loci *in-vivo*. We used comparative gene expression data to find nodes in the conserved eye and head developmental GRN of *D. melanogaster* and *D. mauritiana* that are flexible enough to lead to the observed differences in head shape. Using differential expression dynamics as a readout, we found that higher expression of *pnr* underlies the enlargement of the eye area, a higher number of ommatidia and a narrower face cuticle in *D. mauritiana*. However, using this approach it remains so-far unclear, where the causative mutation lies that leads to variation in *pnr* expression and subsequently to observed changes in adult morphologies. Studying if a gene is differentially expressed due to *cis*- or *trans*-regulatory divergence is a first step to reveal the causative variants (Wittkopp, 2013). Using our genome wide allele specific expression (ASE) dataset, we found that *pnr* itself shows very likely divergent expression due to changes in *trans* in the mid third instar disc (Chapter III). Therefore, Pnr most probably does not represent the evolving locus between *D. melanogaster* and *D. mauritiana*. We do not yet know which upstream factors activate *pnr* expression in the early eye-antennal disc. During early embryonic development *pnr* expression is under the control of Dpp signaling (Ashe et al., 2000; Winick et al., 1993), an interaction that had also been shown for the developing wing imaginal disc (Tomoyasu et al., 2000). Preliminary results that I gained combining our ATAC-seq datasets and transcription factor motif search tools, suggest that Jim, a zinc finger transcription factor or pMad, the transcription factor translating Dpp activity, might be good candidates which are currently tested in our lab.

A genome wide investigation of how the expression of highly connected transcription factors evolve (Chapter III) revealed that if they were not conserved between two species, they were almost exclusively divergent due to *trans*-regulatory changes. This underrepresentation of *cis*-regulatory changes in highly connected transcription factors suggests, that these toolkit genes are not only constraint on a coding sequence level (Halder et al., 1995; King and Wilson, 1975), but also on the level of *cis*-regulatory regions. We could support this hypothesis using our comparative ATAC-seq dataset, by showing that accessible regulatory regions of genes, divergent due to upstream *trans*-changes, are much more conserved on a sequence level than genes that are differentially expressed due to *cis*-regulatory changes. Consistent with this finding is that highly connected genes, have in general a lower log2 fold change between closely related species than genes displaying a low degree (Dr. Torres-Oliva, M.; personal

communication). We conclude, that changes upstream in a developmental GRN are rather due to *trans*-regulatory changes, but that highly pleiotropic factors like Pnr still represent 'flexible nodes' in a conserved GRN, driving phenotypic variation.

It is still an open question if, during evolution of GRNs, we find changes in highly connected nodes, presumably in genes with highly pleiotropic functions, or rather changes in genes located at the endpoints of GRNs. Highly connected nodes can represent so-called 'transcription factor (TF)-hubs' and are defined as transcriptional regulators that impact an exceptionally large number of downstream target genes (MacNeil and Walhout, 2011). Our bioinformatics analysis of target genes suggests that Pnr potentially regulates more than 1000 genes, of which more than 700 were significantly differentially expressed between *D. melanogaster* and *D. mauritiana* during eye-antennal disc development. This high number of putative target genes is consistent with previous studies suggesting that Pnr takes over different functions during eye-and head development. Pnr defines the dorsal lineage of the eye-antennal disc and by this is involved in setting up the dorsal/ventral border (Maurel-Zaffran and Treisman, 2000; Singh et al., 2005; Singh and Choi, 2003). We confirmed this with our lineage tracing line, showing that the complete dorsal part of the eye-antennal disc, including the retinal part, stems from initially *pnr*-expressing cells. Pnr was also suggested to promote head cuticle fate via repression of members of the retinal determination network (Oros et al., 2010). Apart from these many roles during eye and head development, the function of Pnr during *Drosophila* wing development is well characterized (Sato and Saigo, 2000; Tomoyasu et al., 2000). The GATA factor is also crucial for the dorsal/ventral patterning of the embryo (Heitzler et al., 1996; Herranz and Morata, 2001; Winick et al., 1993). Pnr also plays a role in setting up proper sensory bristle patterns (Haenlin et al., 1997; Heitzler et al., 1996) and activates together with Tinman (Tin) *D-mef2*, promoting cardioblast fate in *Drosophila* (Gajewski et al., 1999; Lovato et al., 2015). Given these diverse functions it seems at a first glance counter-intuitive that such a highly pleiotropic factor underlies the evolution of adult morphologies.

A recent example which exemplifies that the wiring and evolution of GRNs is highly context dependent, was elegantly shown with the characterization of variation in another *Drosophila* GRN, namely the one underlying trichome development. These small actin-filled protrusions form in different stages at different positions of the developing fly, for instance on

the larval cuticle, and during pupal stages on developing legs (reviewed in (Arif et al., 2015)). The causative changes that led to repeated loss of these small structures on legs and larval cuticle are though surprisingly different. Whereas changes in the regulatory regions of *shavenbaby* (*svb*) - a key player during trichome formation - lead to loss of these structures in the larva (McGregor et al., 2007; Sucena et al., 2003; Sucena and Stern, 2000), it seems that changes in the *cis*-regulatory regions of the microRNA *miR-92a* cause the appearance of the so-called naked valley on adult *Drosophila* legs (Arif et al., 2013b). Kittelmann and colleagues could show, that even though a similar set of genes governs trichome formation in both structures, some nodes and edges of the underlying GRN differ and that their variation in the wiring can lead to differences in which nodes and (sub-)networks eventually evolve (Kittelmann et al., 2018). Therefore, context-dependent wiring of important developmental nodes might be prevalent.

Our results, that variation in *pnr* expression affects the eye area but also the head cuticle suggests that Pnr is very likely involved in several sub-networks that participate in distinct developmental cascades. Context dependent integration into sub-networks and function might be ensured by spatiotemporal availability of co-factors like Ush, as for instance shown in the wing disc (Fromental-Ramain et al., 2010, 2008). Understanding the role of Pnr and architecture of potential sub-networks will require a more thorough dissection of the effects of Pnr up- or downregulation at distinct time points. The extension of the here used GAL4/UAS system with GAL80 (Jiang et al., 2009; Suster et al., 2004), providing additional temporal control of gene expression could be used in the future to address this question. In summary, with Pnr we found a highly pleiotropic TF-hub acting as a 'flexible node' underlying natural variation in eye size and head shape between *D. melanogaster* and *D. mauritiana*.

Phenotypes that evolved repeatedly, like the re-occurring loss of trichomes, provide a powerful tool to learn more about the evolvability, architecture and robustness of developmental GRNs, by studying if the same nodes or paths are evolving, given the assumption, that the structure of the GRN and therefore the position of a specific node influences where these switches can arise (Stern and Orgogozo, 2008). Categorizing phenotypes and their genetic basis revealed that often evolution at the same loci underlies the evolution of similar traits (Martin and Orgogozo, 2013). Simple traits like trichome patterns have proven to be an excellent model to study which nodes are likely to evolve, since the

underlying GRN is extremely well understood. The *shavenbaby* (*svb*) gene has been proposed to act as a so called 'hot-spot' gene, since repeated changes in the expression of *svb* have been correlated with changes in trichome patterns on cuticles of *Drosophila* larva (McGregor et al., 2007; Sucena et al., 2003; Sucena and Stern, 2000). Stern and Orgogozo proposed that due to its specific position in the GRN, *svb* can act as a switch, turning the development of trichomes simply on or off (Stern and Orgogozo, 2008). Recently the genetic changes underlying repeated loss of pelvic appendages in stickleback fish have been revealed. Freshwater populations of these fish independently lost these structures and mapping data point towards repeated, independent deletions in the enhancer region of the *Pitx1* gene, encoding a homeodomain transcription factor (Chan et al., 2010; Coyle et al., 2007; Cresko et al., 2004; Shapiro et al., 2004). Xie et al. elegantly showed, that the DNA sequence of the *Pitx1* enhancer is exceptionally fragile and therefore prone to break more frequently (Xie et al., 2019), providing a clear mechanistic cause for repeated evolution at the same locus. These examples are cases, where relatively simple traits were studied, and the resulting phenotype is described with a discrete readout, namely loss or gain of a trait. It remains to be shown, which molecular changes underlie the repeated evolution of complex traits - like the here studied trade-off of head structures in different *Drosophila* species. This trade-off between eye size and head width was characterized in several species of the *Drosophila melanogaster* subgroup (Gaspar et al., 2019; Hilbrant et al., 2014; Norry et al., 2000; Posnien et al., 2012). A recent study describes a general inverse resource allocation between the visual system and the olfactory system in more than 60 species within the *Drosophila* genus (Keesey et al., 2019). The authors argued that this trade-off evolved several times independently in the genus (Keesey et al., 2019) and therefore provides an excellent opportunity to study the repeated evolution of a complex trait. Characterization of the trade-off between *D. melanogaster*, *D. simulans* and *D. mauritiana* revealed, that evolution of eye size differences can have two different causes: First, the eye area can be changed by a different number of more or less equally sized single ommatidia. Second, the number of these single facets can be kept stable, but instead the size of the ommatidia can change. A single-nucleotide-polymorphism (SNP) in the regulatory region of *ey* has recently been linked to heterochronic changes in the expression of this master regulator and the authors could functionally validate, that this variant underlies natural variation in ommatidia number of various *D. melanogaster* populations (Ramaekers et al., 2018). Analyses of eye size differences between *D. simulans* and *D. mauritiana* have shown, that the number of

ommatidia does not differ between these two species, but that the latter one develops larger ommatidia, leading to an increased eye ares (Posnien et al., 2012). Preliminary data suggests that changes in the expression of *ocelliless* (*oc*), causes the observed differences in facet size in the two species (Almudi et al. in prep.). Together with our results, this might indicate that in general expression changes in highly pleiotropic factors underlie repeated evolution of this trade-off between eye size and head width, but that the causative molecular mechanisms might be surprisingly different.

7. References

Abzhanov, A., 2004. Bmp4 and Morphological Variation of Beaks in Darwin's Finches. Science 305, 1462–1465. https://doi.org/10.1126/science.1098095

Abzhanov, A., Kuo, W.P., Hartmann, C., Grant, B.R., Grant, P.R., Tabin, C.J., 2006. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. Nature 442, 563–567. https://doi.org/10.1038/nature04843

Adachi, Y., Hauck, B., Clements, J., Kawauchi, H., Kurusu, M., Totani, Y., Kang, Y.Y., Eggert, T., Walldorf, U., Furukubo-Tokunaga, K., Callaerts, P., 2003. Conserved cis-regulatory modules mediate complex neural expression patterns of the eyeless gene in the Drosophila brain. Mechanisms of Development 120, 1113–1126. https://doi.org/10.1016/j.mod.2003.08.007

Agawa, Y., Sarhan, M., Kageyama, Y., Akagi, K., Takai, M., Hashiyama, K., Wada, T., Handa, H., Iwamatsu, A., Hirose, S., Ueda, H., 2007. Drosophila Blimp-1 Is a Transient Transcriptional Repressor That Controls Timing of the Ecdysone-Induced Developmental Pathway. Molecular and Cellular Biology 27, 8739–8747. https://doi.org/10.1128/MCB.01304-07

Akagi, K., Ueda, H., 2011. Regulatory mechanisms of ecdysone-inducible Blimp-1 encoding a transcriptional repressor that is important for the prepupal development in Drosophila: Unique mode of Blimp-1 regulation. Development, Growth & Differentiation 53, 697–703. https://doi.org/10.1111/j.1440-169X.2011.01276.x

Akam, M., 1995. Hox genes and the evolution of diverse body plans. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 349, 313–319. https://doi.org/10.1098/rstb.1995.0119

Akam, M., Averof, M., Castelli-Gair, J., Dawes, R., Falciani, F., Ferrier, D., 1994. The evolving role of Hox genes in arthropods. Dev. Suppl. 209–215.

Akhtar, A., Becker, P.B., 2000. Activation of Transcription through Histone H4 Acetylation by MOF, an Acetyltransferase Essential for Dosage Compensation in Drosophila. Molecular Cell 5, 367–375. https://doi.org/10.1016/S1097-2765(00)80431-1

Alexandre, C.M., Urton, J.R., Jean-Baptiste, K., Huddleston, J., Dorrity, M.W., Cuperus, J.T., Sullivan, A.M., Bemm, F., Jolic, D., Arsovski, A.A., Thompson, A., Nemhauser, J.L., Fields, S., Weigel, D., Bubb, K.L., Queitsch, C., 2018. Complex Relationships between Chromatin Accessibility, Sequence Divergence, and Gene Expression in Arabidopsis thaliana. Molecular Biology and Evolution 35, 837–854. https://doi.org/10.1093/molbev/msx326

Alvarez, M., Schrey, A.W., Richards, C.L., 2015. Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? Mol Ecol 24, 710–725. https://doi.org/10.1111/mec.13055

Alwine, J.C., Kemp, D.J., Stark, G.R., 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proceedings of the National Academy of Sciences 74, 5350–5354. https://doi.org/10.1073/pnas.74.12.5350

Amore, G., Casares, F., 2010. Size matters: the contribution of cell proliferation to the progression of the specification Drosophila eye gene regulatory network. Dev. Biol. 344, 569–577. https://doi.org/10.1016/j.ydbio.2010.06.015

Ando, H., 1962. The comparative embryology of Odonata with special reference to a relic dragonfly Epiophlebia superstes Selys. Sugadaira Biological Laboratory of Tokyo Kyoiku University.

Andō, H., 1962. comparative embryology of Odonata with special reference to a relic dragonfly Epiophlebia superstes Selys.

Ando, H., Haga, K., 1974. Studies on the Pleuropodia of En1bioptera, 9.

Angelini, D.R., Liu, P.Z., Hughes, C.L., Kaufman, T.C., 2005. Hox gene function and interaction in the milkweed bug Oncopeltus fasciatus (Hemiptera). Developmental Biology 287, 440–455. https://doi.org/10.1016/j.ydbio.2005.08.010

Arif, S., Hilbrant, M., Hopfen, C., Almudi, I., Nunes, M.D.S., Posnien, N., Kuncheria, L., Tanaka, K., Mitteroecker, P., Schlötterer, C., McGregor, A.P., 2013a. Genetic and developmental analysis of differences in eye and face morphology between *Drosophila simulans* and *Drosophila mauritiana*: Evolution of eye and face morphology in *Drosophila*. Evolution & Development 15, 257–267. https://doi.org/10.1111/ede.12027

Arif, S., Kittelmann, S., McGregor, A.P., 2015. From shavenbaby to the naked valley: trichome formation as a model for evolutionary developmental biology. Evolution & Development 17, 120–126. https://doi.org/10.1111/ede.12113

Arif, S., Murat, S., Almudi, I., Nunes, M.D.S., Bortolamiol-Becet, D., McGregor, N.S., Currie, J.M.S., Hughes, H., Ronshaugen, M., Sucena, É., Lai, E.C., Schlötterer, C., McGregor, A.P., 2013b. Evolution of mir-92a Underlies Natural Morphological Variation in Drosophila melanogaster. Current Biology 23, 523–528. https://doi.org/10.1016/j.cub.2013.02.018

Arnold, C.D., Gerlach, D., Spies, D., Matts, J.A., Sytnikova, Y.A., Pagani, M., Lau, N.C., Stark, A., 2014. Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution. Nature Genetics 46, 685–692. https://doi.org/10.1038/ng.3009

Arnone, M.I., Davidson, E.H., 1997. The hardwiring of development: organization and function of genomic regulatory systems. Development 124, 1851–1864.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. Nat Genet 25, 25–29. https://doi.org/10.1038/75556

Ashe, H.L., Mannervik, M., Levine, M., 2000. Dpp signaling thresholds in the dorsal ectoderm of the Drosophila embryo. Development 127, 3305–3312.

Auerbach, C., 1936. The development of the legs, wings, and halteres in wild type and some mutant strains of Drosophila melanogaster. Proc. r. Soc. Edinb. 787–815.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research 37, W202–W208. https://doi.org/10.1093/nar/gkp335

Baker, L.R., Weasner, B.M., Nagel, A., Neuman, S.D., Bashirullah, A., Kumar, J.P., 2018. Eyeless/Pax6 initiates eye formation non-autonomously from the peripodial epithelium. Development 145, dev163329. https://doi.org/10.1242/dev.163329

Barabasi, A.-L., Albert, R., 1999. Emergence of Scaling in Random Networks 286, 5.

Bartel, D.P., 2018. Metazoan MicroRNAs. Cell 173, 20–51. https://doi.org/10.1016/j.cell.2018.03.006

Bedford, G.O., 1978. The development of the egg of Didymuria violescens (Phasmatodea: Phasmatidae: Podacanthinae) - embryology and determination of the stage at shich first diapause occurs. Aust. J. Zool. 18, 155–169.

Bennett, R.L., Brown, S.J., Denell, R.E., 1999. Molecular and genetic analysis of the Tribolium Ultrabithorax ortholog, Ultrathorax. Dev. Genes Evol. 209, 608–619.

Bergman, P., Seyedoleslami Esfahani, S., Engström, Y., 2017. Chapter Two - Drosophila as a Model for Human Diseases—Focus on Innate Immunity in Barrier Epithelia, in: Pick, L. (Ed.), Current Topics in Developmental Biology, Fly Models of Human Diseases. Academic Press, pp. 29–81. https://doi.org/10.1016/bs.ctdb.2016.07.002

Bernays, E.A., 1971. The vermiform larva of Schistocerca gregaria (Forskål): Form and activity (Insecta, Orthoptera). Z. Morph. Tiere 70, 183–200. https://doi.org/10.1007/BF00277761

Berridge, M.J., Oschman, J.L., 1972. Transporting Epithelia. Elsevier. https://doi.org/10.1016/B978-0-12-454135-1.X5001-X

Bessa, J., 2002. Combinatorial control of Drosophila eye development by Eyeless, Homothorax, and Teashirt. Genes & Development 16, 2415–2427. https://doi.org/10.1101/gad.1009002

Bird, A.P., Wolffe, A.P., 1999. Methylation-Induced Repression— Belts, Braces, and Chromatin. Cell 99, 451–454. https://doi.org/10.1016/S0092-8674(00)81532-9

Boell, L., Gregorova, S., Forejt, J., Tautz, D., 2011. A comparative assessment of mandible shape in a consomic strain panel of the house mouse (Mus musculus) - implications for epistasis and evolvability of quantitative traits. BMC Evolutionary Biology 11, 309. https://doi.org/10.1186/1471-2148-11-309

Boell, L., Tautz, D., 2011. Micro-evolutionary divergence patterns of mandible shapes in wild house mouse (Mus musculus) populations. BMC Evolutionary Biology 11, 306. https://doi.org/10.1186/1471-2148-11-306

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bonini, N.M., Choi, K.W., 1995. Early decisions in Drosophila eye morphogenesis. Curr. Opin. Genet. Dev. 5, 507–515.

Boyle, E.A., Li, Y.I., Pritchard, J.K., 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169, 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Bozek, M., Cortini, R., Storti, A.E., Unnerstall, U., Gaul, U., Gompel, N., 2019. ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the *Drosophila* blastoderm. Genome Res. 29, 771–783. https://doi.org/10.1101/gr.242362.118

Brennan, C.A., Ashburner, M., Moses, K., 1998. Ecdysone pathway is required for furrow progression in the developing Drosophila eye. Development 125, 2653–2664.

Brown, J.B., Boley, N., Eisman, R., May, G.E., Stoiber, M.H., Duff, M.O., Booth, B.W., Wen, J., Park, S., Suzuki, A.M., Wan, K.H., Yu, C., Zhang, D., Carlson, J.W., Cherbas, L., Eads, B.D., Miller, D., Mockaitis, K., Roberts, J., Davis, C.A., Frise, E., Hammonds, A.S., Olson, S., Shenker, S., Sturgill, D., Samsonova, A.A., Weiszmann, R., Robinson, G., Hernandez, J., Andrews, J., Bickel, P.J., Carninci, P., Cherbas, P., Gingeras, T.R., Hoskins, R.A., Kaufman, T.C., Lai, E.C., Oliver, B., Perrimon, N., Graveley, B.R., Celniker, S.E., 2014. Diversity and dynamics of the Drosophila transcriptome. Nature 512, 393–399. https://doi.org/10.1038/nature12962

Buchberger, Reis, Lu, Posnien, 2019. Cloudy with a Chance of Insights: Context Dependent Gene Regulation and Implications for Evolutionary Studies. Genes 10, 492. https://doi.org/10.3390/genes10070492

Buchon, N., Silverman, N., Cherry, S., 2014. Immunity in *Drosophila melanogaster* — from microbial recognition to whole-organism physiology. Nature Reviews Immunology 14, 796–810. https://doi.org/10.1038/nri3763

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., Greenleaf, W.J., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods 10, 1213–1218. https://doi.org/10.1038/nmeth.2688

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., Greenleaf, W.J., 2015. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 523, 486–490. https://doi.org/10.1038/nature14590

Buffry, A.D., Mendes, C.C., McGregor, A.P., 2016. The Functionality and Evolution of Eukaryotic Transcriptional Enhancers, in: Advances in Genetics. Elsevier, pp. 143–206. https://doi.org/10.1016/bs.adgen.2016.08.004

Bullière, F., 1970. Development of pleuropoda during the embryonic development of Blabera craniifer (insect, Dictyoptera). Arch Anat Microsc Morphol Exp 59, 201–220.

Burgio, G., Baylac, M., Heyer, E., Montagutelli, X., 2009. Genetic analysis of skull shape variation and morphological integration in the mouse using interspecific recombinant congenic strains between C57BL/6 and mice of the mus spretus species. Evolution 63, 2668–2686. https://doi.org/10.1111/j.1558-5646.2009.00737.x

Bustin, S., 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. Journal of Molecular Endocrinology 169–193. https://doi.org/10.1677/jme.0.0250169

Callaerts, P., Halder, G., Gehring, W.J., 1997. *PAX-6* IN DEVELOPMENT AND EVOLUTION. Annu. Rev. Neurosci. 20, 483–532. https://doi.org/10.1146/annurev.neuro.20.1.483

Calleja, M., Moreno, E., Pelaz, S., Morata, G., 1996. Visualization of Gene Expression in Living Adult Drosophila. Science 274, 252–255.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. https://doi.org/10.1186/1471-2105-10-421

Cannavò, E., Koelling, N., Harnett, D., Garfield, D., Casale, F.P., Ciglar, L., Gustafson, H.E., Viales, R.R., Marco-Ferreres, R., Degner, J.F., Zhao, B., Stegle, O., Birney, E., Furlong, E.E.M., 2017. Genetic variants regulating expression levels and isoform diversity during embryogenesis. Nature 541, 402–406. https://doi.org/10.1038/nature20802

Carroll, S.B., 2005. Evolution at Two Levels: On Genes and Form. PLoS Biol 3, e245. https://doi.org/10.1371/journal.pbio.0030245

Carroll, S.B., Grenier, J.K., Weatherbee, S.D., 2001. From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design, 2nd Edition. Wiley-Blackwell.

Casares, F., Almudi, I., 2016. Fast and Furious 800. The Retinal Determination Gene Network in Drosophila, in: Castelli-Gair Hombría, J., Bovolenta, P. (Eds.), Organogenetic Gene Networks: Genetic Control of Organ Formation. Springer International Publishing, Cham, pp. 95–124. https://doi.org/10.1007/978-3-319-42767-6_4

Casasa, S., Moczek, A.P., 2018. Insulin signalling's role in mediating tissue-specific nutritional plasticity and robustness in the horn-polyphenic beetle Onthophagus taurus. Proceedings of the Royal Society B: Biological Sciences 9.

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., Myers, R.M., Petrov, D., Jónsson, B., Schluter, D., Bell, M.A., Kingsley, D.M., 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science 327, 302–305. https://doi.org/10.1126/science.1182213

Chanut, F., Heberlein, U., n.d. Role of the morphogenetic furrow in establishing polarity in the Drosophila eye 10.

Chávez, V.M., Marqués, G., Delbecque, J.P., Kobayashi, K., Hollingsworth, M., Burr, J., Natzle, J.E., O'Connor, M.B., 2000. The Drosophila disembodied gene controls late embryonic morphogenesis and codes for a cytochrome P450 enzyme that regulates embryonic ecdysone levels. Development 127, 4115–4126.

Cheng, H., Wang, Y., Sun, M., 2018. Comparison of Gene Expression Profiles in Nonmodel Eukaryotic Organisms with RNA-Seq, in: Wang, Y., Sun, M. (Eds.), Transcriptome Data Analysis. Springer New York, New York, NY, pp. 3–16. https://doi.org/10.1007/978-1-4939-7710-9_1

Chintapalli, V.R., Wang, J., Herzyk, P., Davies, S.A., Dow, J.A.T., 2013. Data-mining the FlyAtlas online resource to identify core functional motifs across transporting epithelia. BMC Genomics 14, 518. https://doi.org/10.1186/1471-2164-14-518

Chyb, S., Gompel, N., 2013. Wild-type morphology, in: Atlas of Drosophila Morphology. Elsevier, pp. 1–23. https://doi.org/10.1016/B978-0-12-384688-4.00001-8

Ciapponi, L., 2001. Drosophila Fos mediates ERK and JNK signals via distinct phosphorylation sites. Genes & Development 15, 1540–1553. https://doi.org/10.1101/gad.886301

Coolon, J.D., Stevenson, K.R., McManus, C.J., Yang, B., Graveley, B.R., Wittkopp, P.J., 2015. Molecular Mechanisms and Evolutionary Processes Contributing to Accelerated Divergence of Gene Expression on the *Drosophila* X Chromosome. Mol Biol Evol 32, 2605–2615. https://doi.org/10.1093/molbev/msv135

Corrales, M., Rosado, A., Cortini, R., van Arensbergen, J., van Steensel, B., Filion, G.J., 2017. Clustering of *Drosophila* housekeeping promoters facilitates their expression. Genome Res. 27, 1153–1161. https://doi.org/10.1101/gr.211433.116

Cowles, C.R., Hirschhorn, J.N., Altshuler, D., Lander, E.S., 2002. Detection of regulatory variation in mouse genes. Nature Genetics 32, 432–437. https://doi.org/10.1038/ng992

Cox, M.P., Peterson, D.A., Biggs, P.J., 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11, 485. https://doi.org/10.1186/1471-2105-11-485

Coyle, S.M., Huntingford, F.A., Peichel, C.L., 2007. Parallel Evolution of Pitx1 Underlies Pelvic Reduction in Scottish Threespine Stickleback (Gasterosteus aculeatus). Journal of Heredity 98, 581–586. https://doi.org/10.1093/jhered/esm066

Cresko, W.A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., Bell, M.A., Kimmel, C.B., Postlethwait, J.H., 2004. Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. Proceedings of the National Academy of Sciences 101, 6050–6055. https://doi.org/10.1073/pnas.0308479101

Cubadda, Y., Heitzler, P., Ray, R.P., Bourouis, M., Ramain, P., Gelbart, W., Simpson, P., Haenlin, M., 1997. u-shaped encodes a zinc finger protein that regulates the proneural genes achaete and scute during the formation of bristles in Drosophila. Genes & Development 11, 3083–3095. https://doi.org/10.1101/gad.11.22.3083

Cubeñas-Potts, C., Rowley, M.J., Lyu, X., Li, G., Lei, E.P., Corces, V.G., 2017. Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. Nucleic Acids Research 45, 1714–1730. https://doi.org/10.1093/nar/gkw1114

Currea, J.P., Smith, J.L., Theobald, J.C., 2018. Small fruit flies sacrifice temporal acuity to maintain contrast sensitivity. Vision Research 149, 1–8. https://doi.org/10.1016/j.visres.2018.05.007

Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., Shendure, J., 2015. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 348, 910–914. https://doi.org/10.1126/science.aab1601

David, J., McEvey, S., Solignac, M., Tsacas, L., 1989. Drosophila communities on Mauritius and the ecological niche of D. mauritiana (Diptera, Drosophilidae).

Davidson, E.H., 2001. Genomic Regulatory Systems. Academic Press, San Diego.

Davidson, E.H., Levine, M.S., 2008. Properties of developmental gene regulatory networks. Proceedings of the National Academy of Sciences 105, 20063–20066. https://doi.org/10.1073/pnas.0806007105

Davie, K., Jacobs, J., Atkins, M., Potier, D., Christiaens, V., Halder, G., Aerts, S., 2015. Discovery of Transcription Factors and Regulatory Regions Driving In Vivo Tumor Development by ATAC-seq and FAIRE-seq Open Chromatin Profiling. PLOS Genetics 11, e1004994. https://doi.org/10.1371/journal.pgen.1004994

Davis, G.K., Patel, N.H., 2002. Short, Long, and Beyond: Molecular and Embryological Approaches to Insect Segmentation. Annu. Rev. Entomol. 47, 669–699. https://doi.org/10.1146/annurev.ento.47.091201.145251

Dehal, P., Boore, J.L., 2005. Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. PLoS Biology 3, e314. https://doi.org/10.1371/journal.pbio.0030314

Deng, W., Nickle, D.C., Learn, G.H., Maust, B., Mullins, J.I., 2007. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. Bioinformatics 23, 2334–2336. https://doi.org/10.1093/bioinformatics/btm331

Dermitzakis, E.T., 2008. From gene expression to disease risk. Nature Genetics 40, 492–493. https://doi.org/10.1038/ng0508-492

Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C.C., Taylor, J., Burnett, E., Gut, I., Farrall, M., Lathrop, G.M., Abecasis, G.R., Cookson, W.O.C., 2007. A genome-wide association study of global gene expression. Nat Genet 39, 1202–1207. https://doi.org/10.1038/ng2109

Domazet-Lošo, T., Tautz, D., 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. Nature 468, 815–818. https://doi.org/10.1038/nature09632

Domínguez, M., 1999. Hedgehog signalling in the Drosophila eye 126, 9.

Domínguez, M., Casares, F., 2005. Organ specification-growth control connection: new in-sights from the Drosophila eye-antennal disc. Dev. Dyn. 232, 673–684. https://doi.org/10.1002/dvdy.20311

Duboule, D., Dollé, P., 1989. The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. EMBO J. 8, 1497–1505.

Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., Sørensen, M.V., Haddock, S.H.D., Schmidt-Rhaesa, A., Okusu, A.,

Kristensen, R.M., Wheeler, W.C., Martindale, M.Q., Giribet, G., 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452, 745–749. https://doi.org/10.1038/nature06614

Duveau, F., Yuan, D.C., Metzger, B.P.H., Hodgins-Davis, A., Wittkopp, P.J., 2017. Effects of mutation and selection on plasticity of a promoter activity in *Saccharomyces cerevisiae*. Proceedings of the National Academy of Sciences 114, E11218–E11227. https://doi.org/10.1073/pnas.1713960115

Ellegren, H., 2014. Genome sequencing and population genomics in non-model organisms. Trends in Ecology & Evolution 29, 51–63. https://doi.org/10.1016/j.tree.2013.09.008

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S., Mouy, M., Steinthorsdottir, V., Eiriksdottir, G.H., Bjornsdottir, G., Reynisdottir, I., Gudbjartsson, D., Helgadottir, A., Jonasdottir, Aslaug, Jonasdottir, Adalbjorg, Styrkarsdottir, U., Gretarsdottir, S., Magnusson, K.P., Stefansson, H., Fossdal, R., Kristjansson, K., Gislason, H.G., Stefansson, T., Leifsson, B.G., Thorsteinsdottir, U., Lamb, J.R., Gulcher, J.R., Reitman, M.L., Kong, A., Schadt, E.E., Stefansson, K., 2008. Genetics of gene expression and its effect on disease. Nature 452, 423–428. https://doi.org/10.1038/nature06758

Fossett, N., Tevosian, S.G., Gajewski, K., Zhang, Q., Orkin, S.H., Schulz, R.A., 2001. The Friend of GATA proteins U-shaped, FOG-1, and FOG-2 function as negative regulators of blood, heart, and eye development in Drosophila. Proceedings of the National Academy of Sciences 98, 7342–7347. https://doi.org/10.1073/pnas.131215798

Fraulob, M., Beutel, R.G., Machida, R., Pohl, H., 2015. The embryonic development of Stylops ovinae (Strepsiptera, Stylopidae) with emphasis on external morphology. Arthropod Struct Dev 44, 42–68. https://doi.org/10.1016/j.asd.2014.10.001

Freese, N.H., Norris, D.C., Loraine, A.E., 2016. Integrated genome browser: visual analytics platform for genomics. Bioinformatics 32, 2089–2095. https://doi.org/10.1093/bioinformatics/btw069

Fristrom, D., Wilcox, M., Fristrom, J., 1993. The distribution of PS integrins, laminin A and F-actin during key stages in Drosophila wing development. Development 117, 509–523.

Fromental-Ramain, C., Taquet, N., Ramain, P., 2010. Transcriptional interactions between the pannier isoforms and the cofactor U-shaped during neural development in Drosophila. Mechanisms of Development 127, 442–457. https://doi.org/10.1016/j.mod.2010.08.002

Fromental-Ramain, C., Vanolst, L., Delaporte, C., Ramain, P., 2008. pannier encodes two structurally related isoforms that are differentially expressed during Drosophila development and display distinct functions during thorax patterning. Mechanisms of Development 125, 43–57. https://doi.org/10.1016/j.mod.2007.10.008

Fukamizo, T., Kramer, K.J., 1985. Mechanism of chitin hydrolysis by the binary chitinase system in insect moulting fluid. Insect Biochemistry 15, 141–145. https://doi.org/10.1016/0020-1790(85)90001-0

Furlong, E.E.M., Levine, M., 2018. Developmental enhancers and chromosome topology. Science 361, 1341–1345. https://doi.org/10.1126/science.aau0320

Gajewski, K., Fossett, N., Molkentin, J.D., Schulz, R.A., 1999. The zinc finger proteins Pannier and GATA4 function as cardiogenic factors in Drosophila. Development 126, 5679–5688.

Garcia-Fernàndez, J., Holland, P.W.H., 1994. Archetypal organization of the amphioxus Hox gene cluster. Nature 370, 563–566. https://doi.org/10.1038/370563a0

García-García, M.J., Ramain, P., Simpson, P., Modolell, J., 1999. Different contributions of pannier and wingless to the patterning of the dorsal mesothorax of Drosophila. Development 126, 3523–3532.

Garg, A., Srivastava, A., Davis, M.M., O'Keefe, S.L., Chow, L., Bell, J.B., 2007. Antagonizing Scalloped With a Novel Vestigial Construct Reveals an Important Role for Scalloped in *Drosophila melanogaster* Leg, Eye and Optic Lobe Development. Genetics 175, 659–669. https://doi.org/10.1534/genetics.106.063966

Gaspar, P., Arif, S., Sumner-Rooney, L., Kittelmann, M., Stern, D.L., Nunes, M.D.S., McGregor, A.P., 2019. Characterisation of the genetic architecture underlying eye size variation within Drosophila

melanogaster and Drosophila simulans (preprint). Evolutionary Biology. https://doi.org/10.1101/555698

Gautier, M., Yamaguchi, J., Foucaud, J., Loiseau, A., Ausset, A., Facon, B., Gschloessl, B., Lagnel, J., Loire, E., Parrinello, H., Severac, D., Lopez-Roques, C., Donnadieu, C., Manno, M., Berges, H., Gharbi, K., Lawson-Handley, L., Zang, L.-S., Vogel, H., Estoup, A., Prud'homme, B., 2018. The Genomic Basis of Color Pattern Polymorphism in the Harlequin Ladybird. Current Biology 28, 3296-3302.e7. https://doi.org/10.1016/j.cub.2018.08.023

Genissel, A., McIntyre, L.M., Wayne, M.L., Nuzhdin, S.V., 2007. Cis and Trans Regulatory Effects Contribute to Natural Variation in Transcriptome of Drosophila melanogaster. Molecular Biology and Evolution 25, 101–110. https://doi.org/10.1093/molbev/msm247

Gibson, M.C., Schubiger, G., 2001. Drosophila peripodial cells, more than meets the eye? Bioessays 23, 691–697. https://doi.org/10.1002/bies.1098

Gibson, M.C., Schubiger, G., 2000. Peripodial cells regulate proliferation and patterning of Drosophila imaginal discs. Cell 103, 343–350. https://doi.org/10.1016/s0092-8674(00)00125-2

Gilad, Y., Rifkin, S.A., Pritchard, J.K., 2008. Revealing the architecture of gene regulation: the promise of eQTL studies. Trends in Genetics 24, 408–415. https://doi.org/10.1016/j.tig.2008.06.001

Gilbert, D., 2013. Gene-omes built from mRNA seq not genome DNA. 7th annual arthropod genomics symposium.

Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R., Lieb, J.D., 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res. 17, 877–885. https://doi.org/10.1101/gr.5533506

Glaser-Schmitt, A., Parsch, J., 2018. Functional characterization of adaptive variation within a cis-regulatory element influencing Drosophila melanogaster growth. PLOS Biology 16, e2004538. https://doi.org/10.1371/journal.pbio.2004538

Goltsev, Y., Rezende, G.L., Vranizan, K., Lanzaro, G., Valle, D., Levine, M., 2009. Developmental and evolutionary basis for drought tolerance of the Anopheles gambiae embryo. Developmental Biology 330, 462–470. https://doi.org/10.1016/j.ydbio.2009.02.038

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36, 3420–3435. https://doi.org/10.1093/nar/gkn176

Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224. https://doi.org/10.1093/molbev/msp259

Graber, V., 1889. Über den Bau und die phylogenetische Bedeurung der embryonalen Bauchanähnge der Insekten. Biol. Zent. B1 9, 355–363.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29, 644–652. https://doi.org/10.1038/nbt.1883

Graham, A., Papalopulu, N., Krumlauf, R., 1989. The murine and Drosophila homeobox gene complexes have common features of organization and expression. Cell 57, 367–378. https://doi.org/10.1016/0092-8674(89)90912-4

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., Brown, J.B., Cherbas, L., Davis, C.A., Dobin, A., Li, R., Lin, W., Malone, J.H., Mattiuzzo, N.R., Miller, D., Sturgill, D., Tuch, B.B., Zaleski, C., Zhang, D., Blanchette, M., Dudoit, S., Eads, B., Green, R.E., Hammonds, A., Jiang, L., Kapranov, P., Langton, L., Perrimon, N., Sandler, J.E., Wan, K.H., Willingham, A., Zhang, Y., Zou, Y., Andrews, J., Bickel, P.J., Brenner, S.E., Brent, M.R., Cherbas, P., Gingeras, T.R., Hoskins, R.A., Kaufman, T.C., Oliver, B., Celniker, S.E., 2011. The developmental transcriptome of Drosophila melanogaster. Nature 471, 473–479. https://doi.org/10.1038/nature09715

Graze, R.M., McIntyre, L.M., Main, B.J., Wayne, M.L., Nuzhdin, S.V., 2009. Regulatory divergence in Drosophila melanogaster and D. simulans, a genomewide analysis of allele-specific expression. Genetics 183, 547–561, 1SI-21SI. https://doi.org/10.1534/genetics.109.105957

Grellet, P., 1971. Volume variations and DNA content of the nuclei ofScapsipedus marginatus Afz. and Br. (Orthoptera, Gryllidae) during embryogenesis. Wilhelm Roux Arch Entwickl Mech Org 167, 243–265. https://doi.org/10.1007/BF00584252

Grenier, J.K., Carroll, S.B., 2000. Functional evolution of the Ultrabithorax protein. Proceedings of the National Academy of Sciences 97, 704–709. https://doi.org/10.1073/pnas.97.2.704

Grossniklaus, U., Cadigan, K.M., Gehring, W.J., 1994. Three maternal coordinate systems cooperate in the patterning of the Drosophila head. Development 120, 3155–3171.

H. Slifer, E., 1937. The Origin and fate of the Membranes sur- rounding the Grasshopper Egg; together with some Experiments on the Source of the Hatching Enzyme.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8, 1494–1512. https://doi.org/10.1038/nprot.2013.084

Haddrill, P.R., Charlesworth, B., Halligan, D.L., Andolfatto, P., 2005. Patterns of intron sequence evolution in Drosophila are dependent upon length and GC content. Genome Biol. 6, R67. https://doi.org/10.1186/gb-2005-6-8-r67

Haeckl, E., 1879. Anthropogenie, Third Edition. ed. W. Engelmann, Leipzig.

Haeckl, E., 1867. Generelle Morphologie der Organismen. Georg Reimer, Berlin.

Haenlin, M., Cubadda, Y., Blondeau, F., Heitzler, P., Lutz, Y., Simpson, P., Ramain, P., 1997. Transcriptional activity of Pannier is regulated negatively by heterodimerization of the GATA DNA-binding domain with a cofactor encoded by the u-shaped gene of Drosophila. Genes & Development 11, 3096–3108. https://doi.org/10.1101/gad.11.22.3096

Hagan, H.R., 1931. The embryogeny of the polyctenid, Hesperoctenes fumariius Westwood, with the reference to viviparity in insects. J. Morphol. Physiol. 51, 3–115.

Hagen, J.F.D., Mendes, C.C., Tanaka, K.M., Gaspar, P., Kittelmann, M., McGregor, A.P., Nunes, M.D.S., 2018. Tartan underlies the evolution of male genital morphology. bioRxiv. https://doi.org/10.1101/462259

Halder, G., Callaerts, P., Gehring, W., 1995. Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. Science 267, 1788–1792. https://doi.org/10.1126/science.7892602

Hall, B.K., 2003. Evo-Devo: evolutionary developmental mechanisms. Int. J. Dev. Biol. 47, 491–495.

Haynie, J.L., Bryant, P.J., 1986. Development of the eye-antenna imaginal disc and morphogenesis of the adult head inDrosophila melanogaster. J. Exp. Zool. 237, 293–308. https://doi.org/10.1002/jez.1402370302

He, X., Zhang, J., 2006. Why Do Hubs Tend to Be Essential in Protein Networks? PLoS Genetics 2, 9.

Heberlein, U., Moses, K., 1995. Mechanisms of Drosophila retinal morphogenesis: the virtues of being progressive. Cell 81, 987–990. https://doi.org/10.1016/s0092-8674(05)80003-0

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., Glass, C.K., 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell 38, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Heitzler, P., Haenlin, M., Ramain, P., Calleja, M., Simpson, P., 1996. A genetic analysis of pannier, a gene necessary for viability of dorsal tissues and bristle positioning in Drosophila. Genetics 143, 1271–1286.

Heming, B.S., 1993. Origin and fate of pleuropodia in embryos of Neoheegeria verbasci (Osborn) (Thysanoptera: Phlaeothripidae). Journal of Pure and Applied Zoology 4, 205–223.

Henikoff, S., Ahmad, K., Malik, H.S., 2001. The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. Science 293, 1098–1102. https://doi.org/10.1126/science.1062939

Herranz, H., Morata, G., 2001. The functions of pannier during Drosophila embryogenesis. Development 128, 4837–4846.

Herrmann, C., Van de Sande, B., Potier, D., Aerts, S., 2012. i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. Nucleic Acids Research 40, e114–e114. https://doi.org/10.1093/nar/gks543

Hilbrant, M., Almudi, I., Leite, D.J., Kuncheria, L., Posnien, N., Nunes, M.D., McGregor, A.P., 2014. Sexual dimorphism and natural variation within and among species in the Drosophilaretinal mosaic. BMC Evol Biol 14, 240. https://doi.org/10.1186/s12862-014-0240-x

Hill, R.E., Favor, J., Hogan, B.L.M., Ton, C.C.T., Saunders, G.F., Hanson, I.M., Prosser, J., Jordan, T., Hastie, N.D., Heyningen, V. van, 1991. Mouse Small eye results from mutations in a paired-like homeobox-containing gene. Nature 354, 522. https://doi.org/10.1038/354522a0

Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., Hillman-Jackson, J., Kuhn, R.M., Pedersen, J.S., Pohl, A., Raney, B.J., Rosenbloom, K.R., Siepel, A., Smith, K.E., Sugnet, C.W., Sultan-Qurraie, A., Thomas, D.J., Trumbower, H., Weber, R.J., Weirauch, M., Zweig, A.S., Haussler, D., Kent, W.J., 2006. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res 34, D590–D598. https://doi.org/10.1093/nar/gkj144

Hoekstra, H.E., Coyne, J.A., 2007. THE LOCUS OF EVOLUTION: EVO DEVO AND THE GENETICS OF ADAPTATION: THE LOCUS OF EVOLUTION. Evolution 61, 995–1016. https://doi.org/10.1111/j.1558-5646.2007.00105.x

Hoekstra, H.E., Hirschmann, R.J., Bundey, R.A., Insel, P.A., Crossland, J.P., 2006. A Single Amino Acid Mutation Contributes to Adaptive Beach Mouse Color Pattern. Science 313, 101–104. https://doi.org/10.1126/science.1126121

Hogenkamp, D.G., Arakane, Y., Kramer, K.J., Muthukrishnan, S., Beeman, R.W., 2008. Characterization and expression of the beta-N-acetylhexosaminidase gene family of Tribolium castaneum. Insect Biochem. Mol. Biol. 38, 478–489. https://doi.org/10.1016/j.ibmb.2007.08.002

Homem, C.C.F., Knoblich, J.A., 2012. Drosophila neuroblasts: a model for stem cell biology. Development 139, 4297–4310. https://doi.org/10.1242/dev.080515

Huang, X., Warren, J.T., Buchanan, J., Gilbert, L.I., Scott, M.P., 2007. Drosophila Niemann-Pick type C-2 genes control sterol homeostasis and steroid biosynthesis: a model of human neurodegenerative disease. Development 134, 3733–3742. https://doi.org/10.1242/dev.004572

Hughes, A.E.O., Enright, J.M., Myers, C.A., Shen, S.Q., Corbo, J.C., 2017. Cell Type-Specific Epigenomic Analysis Reveals a Uniquely Closed Chromatin Architecture in Mouse Rod Photoreceptors. Sci Rep 7, 43184. https://doi.org/10.1038/srep43184

Hughes, C.L., Kaufman, T.C., 2002. Hox genes and the evolution of the arthropod body plan. Evol. Dev. 4, 459–499.

Hussey, P.B., 1926. Studies on the Pleuropodia of Belostoma Flumineum Say and Ranatra Fusca Palisot de Beauvois, wie a discussion of these organs in other insects. Entomol. Am. 1–82.

Imrichová, H., Hulselmans, G., Kalender Atak, Z., Potier, D., Aerts, S., 2015. i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. Nucleic Acids Res 43, W57–W64. https://doi.org/10.1093/nar/gkv395

Jacobs, C.G.C., Braak, N., Lamers, G.E.M., van der Zee, M., 2015. Elucidation of the serosal cuticle machinery in the beetle Tribolium by RNA sequencing and functional analysis of Knickkopf1, Retroactive and Laccase2. Insect Biochem. Mol. Biol. 60, 7–12. https://doi.org/10.1016/j.ibmb.2015.02.014

Jacobs, C.G.C., Rezende, G.L., Lamers, G.E.M., van der Zee, M., 2013. The extraembryonic serosa protects the insect egg against desiccation. Proceedings of the Royal Society B: Biological Sciences 280, 20131082–20131082. https://doi.org/10.1098/rspb.2013.1082

Jacobs, C.G.C., Spaink, H.P., van der Zee, M., 2014. The extraembryonic serosa is a frontier epithelium providing the insect egg with a full-range innate immune response. eLife 3, e04111. https://doi.org/10.7554/eLife.04111

Jarman, A.P., Grell, E.H., Ackerman, L., Jan, L.Y., Jan, Y.N., 1994. atonal is the proneural gene for Drosophila photoreceptors. Nature 369, 398–400. https://doi.org/10.1038/369398a0

Jenny, A., 2010. Planar Cell Polarity Signaling in the Drosophila Eye, in: Current Topics in Developmental Biology. Elsevier, pp. 189–227. https://doi.org/10.1016/B978-0-12-385044-7.00007-2

Jeong, S., Rokas, A., Carroll, S.B., 2006. Regulation of Body Pigmentation by the Abdominal-B Hox Protein and Its Gain and Loss in Drosophila Evolution. Cell 125, 1387–1399. https://doi.org/10.1016/j.cell.2006.04.043

Jia, Z., Xu, S., 2007. Mapping Quantitative Trait Loci for Expression Abundance. Genetics 176, 611–623. https://doi.org/10.1534/genetics.106.065599

Jiang, F., Frey, B.R., Evans, M.L., Friel, J.C., Hopper, J.E., 2009. Gene Activation by Dissociation of an Inhibitor from a Transcriptional Activation Domain. Molecular and Cellular Biology 29, 5604–5610. https://doi.org/10.1128/MCB.00632-09

Johnson, D.S., Mortazavi, A., Myers, R.M., Wold, B., 2007. Genome-Wide Mapping of in Vivo Protein-DNA Interactions. Science 316, 1497–1502. https://doi.org/10.1126/science.1141319

Johnston, D.S., Nüsslein-Volhard, C., 1992. The origin of pattern and polarity in the Drosophila embryo. Cell 68, 201–219. https://doi.org/10.1016/0092-8674(92)90466-P

Jones, B.M., 1956. ENDOCRINE ACTIVITY DURING INSECT EMBRYO- GENESIS. CONTROL OF EVENTS IN DEVELOPMENT FOLLOWING THE EMBRYONIC MOULT (LOCUSTA MIGRATORIA AND LOCUST AN A PARDALINA, ORTHOPTERA). J Exp Biol 33, 685–696.

Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E.H., Birney, E., Furlong, E.E.M., 2012. A Transcription Factor Collective Defines Cardiac Cell Fate and Reflects Lineage History. Cell 148, 473–486. https://doi.org/10.1016/j.cell.2012.01.030

Käll, L., Krogh, A., Sonnhammer, E.L.L., 2007. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Res. 35, W429-432. https://doi.org/10.1093/nar/gkm256

Kamiya, A., Ando, H., 1985. External morphogenesis of the embryo of Ascalaphus ramburi (Neuroptera, Ascalaphidae)., in: Recent Advances in Insect Embryology in Japan. ISEBU Co.Led., pp. 203–213.

Karcavich, R.E., 2005. Generating neuronal diversity in the Drosophila central nervous system: a view from the ganglion mother cells. Dev. Dyn. 232, 609–616. https://doi.org/10.1002/dvdy.20273

Keesey, I.W., Grabe, V., Gruber, L., Koerte, S., Obiero, G.F., Bolton, G., Khallaf, M.A., Kunert, G., Lavista-Llanos, S., Valenzano, D.R., Rybak, J., Barrett, B.A., Knaden, M., Hansson, B.S., 2019. Inverse resource allocation between vision and olfaction across the genus Drosophila. Nat Commun 10, 1162. https://doi.org/10.1038/s41467-019-09087-z

Kent, W.J., 2002. BLAT--the BLAST-like alignment tool. Genome Res. 12, 656–664. https://doi.org/10.1101/gr.229202

Kenyon, K.L., Ranade, S.S., Curtiss, J., Mlodzik, M., Pignoni, F., 2003. Coordinating Proliferation and Tissue Specification to Promote Regional Identity in the Drosophila Head. Developmental Cell 5, 403–414. https://doi.org/10.1016/S1534-5807(03)00243-0

Khaitovich, P., Enard, W., Lachmann, M., Pääbo, S., 2006. Evolution of primate gene expression. Nature Reviews Genetics 7, 693–702. https://doi.org/10.1038/nrg1940

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., Baranasic, D., Arenillas, D.J., Sandelin, A., Vandepoele, K., Lenhard, B., Ballester, B., Wasserman, W.W., Parcy, F., Mathelier, A., 2018. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Research 46, D260–D266. https://doi.org/10.1093/nar/gkx1126

Khoueiry, P., Girardot, C., Ciglar, L., Peng, P.-C., Gustafson, E.H., Sinha, S., Furlong, E.E., 2017. Uncoupling evolutionary changes in DNA sequence, transcription factor occupancy and enhancer activity. eLife 6. https://doi.org/10.7554/eLife.28440

Kim, A.-R., Choi, E.-B., Kim, M.-Y., Choi, K.-W., 2017. Angiotensin-converting enzyme Ance is cooperatively regulated by Mad and Pannier in Drosophila imaginal discs. Sci Rep 7, 13174. https://doi.org/10.1038/s41598-017-13487-w

King, M., Wilson, A., 1975. Evolution at two levels in humans and chimpanzees. Science 188, 107–116. https://doi.org/10.1126/science.1090005

Kittelmann, McGregor, 2019. Modulation and Evolution of Animal Development through microRNA Regulation of Gene Expression. Genes 10, 321. https://doi.org/10.3390/genes10040321

Kittelmann, S., Buffry, A.D., Franke, F.A., Almudi, I., Yoth, M., Sabaris, G., Couso, J.P., Nunes, M.D.S., 2018. Gene regulatory network architecture in different developmental contexts influences the genetic basis of morphological evolution. PLoS Genet 14(5), 21. https://doi.org/10.1371/journal.pgen.1007375

Kjer, K.M., Simon, C., Yavorskaya, M., Beutel, R.G., 2016. Progress, pitfalls and parallel universes: a history of insect phylogenetics. J R Soc Interface 13. https://doi.org/10.1098/rsif.2016.0363

Klingenberg, C.P., 2011. MorphoJ: an integrated software package for geometric morphometrics: COMPUTER PROGRAM NOTE. Molecular Ecology Resources 11, 353–357. https://doi.org/10.1111/j.1755-0998.2010.02924.x

Kobayashi, Y., Ando, H., 1990. Early embryonic development and external features of developing embryos of the caddisfly, Nemotaulius admorsus (trichoptera: Limnephilidae). J. Morphol. 203, 69–85. https://doi.org/10.1002/jmor.1052030108

Kobayashi, Y., Suzuki, H., Ohba, N., 2003. Development of pleuropodia in the embryo of the glowworm Rhagophthalmus ohbai (Rhagophthalmidae, Coleoptera, Insecta), with comments on their probable function.

Konopová, B., Buchberger, E., Crisp, A., 2019. Transcriptomics supports that pleuropodia of insect embryos function in degradation of the serosal cuticle to enable hatching (preprint). Developmental Biology. https://doi.org/10.1101/584029

Konopová, B., Zrzavý, J., 2005. Ultrastructure, development, and homology of insect embryonic cuticles. J. Morphol. 264, 339–362. https://doi.org/10.1002/jmor.10338

Kopp, F., Mendell, J.T., 2018. Functional Classification and Experimental Dissection of Long Noncoding RNAs. Cell 172, 393–407. https://doi.org/10.1016/j.cell.2018.01.011

Koschützki, D., Schreiber, F., 2008. Centrality Analysis Methods for Biological Networks and Their Application to Gene Regulatory Networks. Gene Regul Syst Bio 2, GRSB.S702. https://doi.org/10.4137/GRSB.S702

Kouzarides, T., 2007. Chromatin Modifications and Their Function. Cell 128, 693–705. https://doi.org/10.1016/j.cell.2007.02.005

Kratochwil, C.F., Liang, Y., Gerwin, J., Woltering, J.M., Urban, S., Henning, F., Machado-Schiaffino, G., Hulsey, C.D., Meyer, A., 2018. Agouti-related peptide 2 facilitates convergent evolution of stripe patterns across cichlid fish radiations. Science 362, 457–460. https://doi.org/10.1126/science.aao6809

Krishnan, J., Rohner, N., 2016. Cavefish and the basis for eye loss. Phil.Trans. R. Soc. B 10. http://dx.doi.org/10.1098/rstb.2015.0487

Kumar, J.P., 2018. The fly eye: Through the looking glass: Future of The Fly Eye. Developmental Dynamics 247, 111–123. https://doi.org/10.1002/dvdy.24585

Kumar, J.P., 2009. The molecular circuitry governing retinal determination. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1789, 306–314. https://doi.org/10.1016/j.bbagrm.2008.10.001

Kumar, J.P., Jamal, T., Doetsch, A., Turner, F.R., Duffy, J.B., 2004. CREB Binding Protein Functions During Successive Stages of Eye Development in Drosophila. Genetics 168, 877–893. https://doi.org/10.1534/genetics.104.029850

Kumar, J.P., Moses, K., 2001. The EGF receptor and notch signaling pathways control the initiation of the morphogenetic furrow during Drosophila eye development. Development 128, 2689–2697.

Kvon, E.Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., Stark, A., 2014. Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature 512, 91–95. https://doi.org/10.1038/nature13395

Lambiase, S., Grigolo, A., Morbini, P., 2003. Ontogenesis of pleuropodia in different species of Blattaria (Insecta): A comparative study. https://doi.org/10.1080/11250000309356518

Landry, C.R., Lemos, B., Rifkin, S.A., Dickinson, W.J., Hartl, D.L., 2007. Genetic Properties Influencing the Evolvability of Gene Expression. Science 317, 118–121. https://doi.org/10.1126/science.1140247

Landry, C.R., Wittkopp, P.J., Taubes, C.H., Ranz, J.M., Clark, A.G., Hartl, D.L., 2005. Compensatory *cis-trans* Evolution and the Dysregulation of Gene Expression in Interspecific Hybrids of Drosophila. Genetics 171, 1813–1822. https://doi.org/10.1534/genetics.105.047449

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25. https://doi.org/10.1186/gb-2009-10-3-r25

Larink, O., 1983. Embryonic and postembryonic development of Machilidae and Lepismatidae (Insecta: Archaeognatha et Zygentoma).

Laumer, C.E., Fernández, R., Lemer, S., Combosch, D., Kocot, K.M., Riesgo, A., Andrade, S.C.S., Sterrer, W., Sørensen, M.V., Giribet, G., 2019. Revisiting metazoan phylogeny with genomic sampling of all phyla. Proc. R. Soc. B 286, 20190831. https://doi.org/10.1098/rspb.2019.0831

Lawrence, M., Daujat, S., Schneider, R., 2016. Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends in Genetics 32, 42–56. https://doi.org/10.1016/j.tig.2015.10.007

Lee, M.S.Y., Palci, A., 2015. Morphological Phylogenetics in the Genomic Age. Current Biology 25, R922–R929. https://doi.org/10.1016/j.cub.2015.07.009

Lemaitre, B., Hoffmann, J., 2007. The host defense of Drosophila melanogaster. Annu. Rev. Immunol. 25, 697–743. https://doi.org/10.1146/annurev.immunol.25.022106.141615

Lemos, B., Meiklejohn, C.D., Cáceres, M., Hartl, D.L., 2005. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. Evolution 59, 126–137.

Lenaerts, C., Van Wielendaele, P., Peeters, P., Vanden Broeck, J., Marchal, E., 2016. Ecdysteroid signalling components in metamorphosis and development of the desert locust, Schistocerca gregaria. Insect Biochem. Mol. Biol. 75, 10–23. https://doi.org/10.1016/j.ibmb.2016.05.003

Lewis, D.L., DeCamillis, M., Bennett, R.L., 2000. Distinct roles of the homeotic genes Ubx and abd-A in beetle embryonic abdominal appendage development. PNAS 97, 4504–4509. https://doi.org/10.1073/pnas.97.9.4504

Li, D., Zhang, Jianqin, Wang, Y., Liu, X., Ma, E., Sun, Y., Li, S., Zhu, K.Y., Zhang, Jianzhen, 2015. Two chitinase 5 genes from Locusta migratoria: molecular characteristics and functional differentiation. Insect Biochem. Mol. Biol. 58, 46–54. https://doi.org/10.1016/j.ibmb.2015.01.004

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, S., Li, Y., Carthew, R.W., Lai, Z.-C., 1997. Photoreceptor Cell Differentiation Requires Regulated Proteolysis of the Transcriptional Repressor Tramtrack. Cell 90, 469–478. https://doi.org/10.1016/S0092-8674(00)80507-3

Liang, P., Pardee, A.B., 2003. Analysing differential gene expression in cancer. Nature Reviews Cancer 3, 869. https://doi.org/10.1038/nrc1214

Lim, J., Choi, K.-W., 2004. Drosophila eye disc margin is a center for organizing long-range planar polarity. genesis 39, 26–37. https://doi.org/10.1002/gene.20022

Liu, H.-W., Wang, L.-L., Tang, X., Dong, Z.-M., Guo, P.-C., Zhao, D.-C., Xia, Q.-Y., Zhao, P., 2018. Proteomic analysis of Bombyx mori molting fluid: Insights into the molting process. J Proteomics 173, 115–125. https://doi.org/10.1016/j.jprot.2017.11.027

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–408. https://doi.org/10.1006/meth.2001.1262

Locke, M., Krishnan, N., 1973. The formation of the ecdysial droplets and the ecdysial membrane in an insect. Tissue Cell 5, 441–450.

Lopes, C.S., Casares, F., 2010. hth maintains the pool of eye progenitors and its downregulation by Dpp and Hh couples retinal fate acquisition with cell cycle exit. Developmental Biology 339, 78–88. https://doi.org/10.1016/j.ydbio.2009.12.020

Losos, J.B. (Ed.), 2014. The Princeton guide to evolution. Princeton University Press, Princeton ; Oxford.

Louvet, J.-P., 1983. Ultrastructure du pleuropode chez l'embryon du hanneton Rhizotrogus majalis razoum (Coleoptera : Melolonthidae). International Journal of Insect Morphology and Embryology 12, 97–117. https://doi.org/10.1016/0020-7322(83)90003-X

Louvet, J.P., 1975. Premieres observations sur l'ultrastructure du pleuropode chez le criquet migrateur. Comptes rendus hebdomadaires des seances. Serie D. Sciences naturelles.

Louvet, J.-P., 1973. L'ultrastructure du pleuropode et son ontogenèse, chez l'embryon du phasme arausius morosus Br. I. - Étude du pleuropode de l'embryon agé. Ann 948 Sci Nat Zool 12, 525–594.

Lovato, T.L., Sensibaugh, C.A., Swingle, K.L., Martinez, M.M., Cripps, R.M., 2015. The Drosophila Transcription Factors Tinman and Pannier Activate and Collaborate with Myocyte Enhancer Factor-2 to Promote Heart Cell Fate. PLOS ONE 10, e0132965. https://doi.org/10.1371/journal.pone.0132965

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15. https://doi.org/10.1186/s13059-014-0550-8

Lübbe, A., Schaffner, W., 1985. Perspectives from Molecular Biology 5.

Ludwig, M.Z., Bergman, C., Patel, N.H., Kreitman, M., 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. Nature 403, 564–567. https://doi.org/10.1038/35000615

Luscombe, N.M., Madan Babu, M., Yu, H., Snyder, M., Teichmann, S.A., Gerstein, M., 2004. Genomic analysis of regulatory network dynamics reveals large topological changes. Nature 431, 308–312. https://doi.org/10.1038/nature02782

Machida, R., 1981. External features of embryonic development of a jumping bristletail, Pedetontus unimaculatus Machida (Insecta, Thysanura, Machilidae). J. Morphol. 168, 339–355. https://doi.org/10.1002/jmor.1051680310

Machida, R., Tojo, K., Tsutsumi, T., Uchifune, T., Klass, K.-D., Picker, M.D., Pretorius, L., 2004. Embryonic development of heel-walkers: Reference to some prerevolutionary stages (Insecta: Mantophasmatodea).

Mack, K., Phifer-Rixey, M., Harr, B., Nachman, M., 2019. Gene Expression Networks Across Multiple Tissues Are Associated with Rates of Molecular Evolution in Wild House Mice. Genes 10, 225. https://doi.org/10.3390/genes10030225

Mackay, T.F.C., 2001. Quantitative trait loci in Drosophila. Nat Rev Genet 2, 11–20. https://doi.org/10.1038/35047544

Mackay, T.F.C., Stone, E.A., Ayroles, J.F., 2009. The genetics of quantitative traits: challenges and prospects. Nat Rev Genet 10, 565–577. https://doi.org/10.1038/nrg2612

MacNeil, L.T., Walhout, A.J.M., 2011. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome Research 21, 645–657. https://doi.org/10.1101/gr.097378.109

Maga, A.M., Navarro, N., Cunningham, M.L., Cox, T.C., 2015. Quantitative trait loci affecting the 3D skull shape and size in mouse and prioritization of candidate genes in-silico. Front. Physiol. 6. https://doi.org/10.3389/fphys.2015.00092

Magri, M.S., Domínguez-Cejudo, M.A., Casares, F., 2018. Wnt controls the medial–lateral subdivision of the *Drosophila* head. Biology Letters 14, 20180258. https://doi.org/10.1098/rsbl.2018.0258

Mähler, N., Wang, J., Terebieniec, B.K., Ingvarsson, P.K., Street, N.R., Hvidsten, T.R., 2017. Gene co-expression network connectivity is an important determinant of selective constraint. PLoS Genet 13, e1006402. https://doi.org/10.1371/journal.pgen.1006402

Marchal, E., Badisco, L., Verlinden, H., Vandersmissen, T., Van Soest, S., Van Wielendaele, P., Vanden Broeck, J., 2011. Role of the Halloween genes, Spook and Phantom in ecdysteroidogenesis in the desert locust, Schistocerca gregaria. J. Insect Physiol. 57, 1240–1248. https://doi.org/10.1016/j.jinsphys.2011.05.009

Marchal, E., Verlinden, H., Badisco, L., Van Wielendaele, P., Vanden Broeck, J., 2012. RNAi-mediated knockdown of Shade negatively affects ecdysone-20-hydroxylation in the desert locust, Schistocerca gregaria. J. Insect Physiol. 58, 890–896. https://doi.org/10.1016/j.jinsphys.2012.03.013

Marlétaz, F., Firbas, P.N., Maeso, I., Tena, J.J., Bogdanovic, O., Perry, M., Wyatt, C.D.R., de la Calle-Mustienes, E., Bertrand, S., Burguera, D., Acemel, R.D., van Heeringen, S.J., Naranjo, S., Herrera-Ubeda, C., Skvortsova, K., Jimenez-Gancedo, S., Aldea, D., Marquez, Y., Buono, L., Kozmikova, I., Permanyer, J., Louis, A., Albuixech-Crespo, B., Le Petillon, Y., Leon, A., Subirana, L., Balwierz, P.J., Duckett, P.E., Farahani, E., Aury, J.-M., Mangenot, S., Wincker, P., Albalat, R., Benito-Gutiérrez, È., Cañestro, C., Castro, F., D'Aniello, S., Ferrier, D.E.K., Huang, S., Laudet, V., Marais, G.A.B., Pontarotti, P., Schubert, M., Seitz, H., Somorjai, I., Takahashi, T., Mirabeau, O., Xu, A., Yu, J.-K., Carninci, P., Martinez-Morales, J.R., Crollius, H.R., Kozmik, Z., Weirauch, M.T., Garcia-Fernàndez, J., Lister, R., Lenhard, B., Holland, P.W.H., Escriva, H., Gómez-Skarmeta, J.L., Irimia, M., 2018. Amphioxus functional genomics and the origins of vertebrate gene regulation. Nature 564, 64–70. https://doi.org/10.1038/s41586-018-0734-6

Martin, A., Orgogozo, V., 2013. THE LOCI OF REPEATED EVOLUTION: A CATALOG OF GENETIC HOTSPOTS OF PHENOTYPIC VARIATION: PERSPECTIVE. Evolution n/a-n/a. https://doi.org/10.1111/evo.12081

Mashimo, Y., Beutel, R.G., Dallai, R., Lee, C.-Y., Machida, R., 2014. Embryonic development of Zoraptera with special reference to external morphology, and its phylogenetic implications (Insecta). J. Morphol. 275, 295–312. https://doi.org/10.1002/jmor.20215

Maurel-Zaffran, C., Treisman, J.E., 2000. pannier acts upstream of wingless to direct dorsal eye disc development in Drosophila. Development 127, 1007–1016.

McGinnis, W., Krumlauf, R., 1992. Homeobox genes and axial patterning. Cell 68, 283–302. https://doi.org/10.1016/0092-8674(92)90471-N

McGregor, A.P., Orgogozo, V., Delon, I., Zanet, J., Srinivasan, D.G., Payre, F., Stern, D.L., 2007. Morphological evolution through multiple cis-regulatory mutations at a single gene. Nature 448, 587–590. https://doi.org/10.1038/nature05988

McKay, D.J., Lieb, J.D., 2013. A Common Set of DNA Regulatory Elements Shapes Drosophila Appendages. Developmental Cell 27, 306–318. https://doi.org/10.1016/j.devcel.2013.10.009

McManus, C.J., Coolon, J.D., Duff, M.O., Eipper-Mains, J., Graveley, B.R., Wittkopp, P.J., 2010. Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Research 20, 816–825. https://doi.org/10.1101/gr.102491.109

Merzendorfer, H., 2013. Chitin synthesis inhibitors: old molecules and new developments. Insect Science 20, 121–138. https://doi.org/10.1111/j.1744-7917.2012.01535.x

Metzger, B.P.H., Wittkopp, P.J., Coolon, Joseph.D., 2017. Evolutionary Dynamics of Regulatory Changes Underlying Gene Expression Divergence among Saccharomyces Species. Genome Biology and Evolution 9, 843–854. https://doi.org/10.1093/gbe/evx035

Miller, A., 1940. Embryonic Membranes, Yolk Cells, and Morphogenesis of the Stonefly Pteronarcys Proteus Newman, (Plecoptera: Pteronarcidae). Ann Entomol Soc Am 33, 437–477. https://doi.org/10.1093/aesa/33.2.437

Milner, M.J., Bleasby, A.J., Kelly, S.L., 1984. The role of the peripodial membrane of leg and wing imaginal discs ofDrosophila melanogaster during evagination and differentiation in vitro. Wilhelm Roux' Archiv 193, 180–186. https://doi.org/10.1007/BF00848893

Mishra, A.K., Bargmann, B.O.R., Tsachaki, M., Fritsch, C., Sprecher, S.G., 2016. Functional genomics identifies regulators of the phototransduction machinery in the Drosophila larval eye and adult ocelli. Developmental Biology 410, 164–177. https://doi.org/10.1016/j.ydbio.2015.12.026

Miyakawa, K., 1979. Embryology of the dobsonfly, Protohermes grandis Thunberg (Megaloptera: Corydalidae). I. Changes in external form of the embryo during development. Konchu. Kontyu.

Mlodzik, M., 2002. Planar cell polarization: do the same mechanisms regulate Drosophila tissue polarity and vertebrate gastrulation? Trends in Genetics 18, 564–571. https://doi.org/10.1016/S0168-9525(02)02770-1

Mohamed, A.A., Zhang, L., Dorrah, M.A., Elmogy, M., Yousef, H.A., Bassal, T.T.M., Duvic, B., 2016. Molecular characterization of a c-type lysozyme from the desert locust, Schistocerca gregaria (Orthoptera: Acrididae). Dev. Comp. Immunol. 61, 60–69. https://doi.org/10.1016/j.dci.2016.03.018

Moriyama, Y., Koshiba-Takeuchi, K., 2018. Significance of whole-genome duplications on the emergence of evolutionary novelties. Briefings in Functional Genomics 17, 329–338. https://doi.org/10.1093/bfgp/ely007

Muratoglu, S., Garratt, B., Hyman, K., Gajewski, K., Schulz, R.A., Fossett, N., 2006. Regulation of Drosophila friend of GATA gene, u-shaped, during hematopoiesis: a direct role for serpent and lozenge. Dev. Biol. 296, 561–579. https://doi.org/10.1016/j.ydbio.2006.04.455

Neto, M., Naval-Sánchez, M., Potier, D., Pereira, P.S., Geerts, D., Aerts, S., Casares, F., 2017. Nuclear receptors connect progenitor transcription factors to cell cycle control. Sci Rep 7, 4845. https://doi.org/10.1038/s41598-017-04936-7

Nijhout, F.J., 1998. Insect Hormones.

Niven, J.E., Laughlin, S.B., 2008. Energy limitation as a selective pressure on the evolution of sensory systems. J. Exp. Biol. 211, 1792–1804. https://doi.org/10.1242/jeb.017574

Niwa, R., Matsuda, T., Yoshiyama, T., Namiki, T., Mita, K., Fujimoto, Y., Kataoka, H., 2004. CYP306A1, a cytochrome P450 enzyme, is essential for ecdysteroid biosynthesis in the prothoracic glands of Bombyx and Drosophila. J. Biol. Chem. 279, 35942–35949. https://doi.org/10.1074/jbc.M404514200

Niwa, R., Niwa, Y.S., 2014. Enzymes for ecdysteroid biosynthesis: their biological functions in insects and beyond. Biosci. Biotechnol. Biochem. 78, 1283–1292. https://doi.org/10.1080/09168451.2014.942250

Noh, M.Y., Muthukrishnan, S., Kramer, K.J., Arakane, Y., 2018. A chitinase with two catalytic domains is required for organization of the cuticular extracellular matrix of a beetle. PLoS Genet. 14, e1007307. https://doi.org/10.1371/journal.pgen.1007307

Norling, U., 1982. Structure and ontogeny of the lateral abdominal gills and the caudal gills in Euphaeidae (Odonata: Zygoptera) larvae. Zoologische Jahrbucher. Abteilung fur Anatomie und Ontogenie der Tiere.

Norry, F.M., Gomez, F.H., 2017. Quantitative Trait Loci and Antagonistic Associations for Two Developmentally Related Traits in the *Drosophila* Head. Journal of Insect Science 17, 19. https://doi.org/10.1093/jisesa/iew115

Norry, F.M., Vilardi, J.C., Hasson, E., 2000. Negative genetic correlation between traits of the Drosophila head, and interspecific divergence in head shape. Heredity (Edinb) 85 (Pt 2), 177–183. https://doi.org/10.1046/j.1365-2540.2000.00735.x

Novak, V.J.A., Zambre, S.K., 1974. To the problem of structure and function of pleuropodia in Schistocerca gregaria Forskal embryos. Zoologische Jahrbücher.

Oakley, T.H., Wolfe, J.M., Lindgren, A.R., Zaharoff, A.K., 2013. Phylotranscriptomics to Bring the Understudied into the Fold: Monophyletic Ostracoda, Fossil Placement, and Pancrustacean Phylogeny. Molecular Biology and Evolution 30, 215–233. https://doi.org/10.1093/molbev/mss216

Oros, S.M., Tare, M., Kango-Singh, M., Singh, A., 2010. Dorsal eye selector pannier (pnr) suppresses the eye fate to define dorsal margin of the Drosophila eye. Developmental Biology 346, 258–271. https://doi.org/10.1016/j.ydbio.2010.07.030

Ou, Q., Zeng, J., Yamanaka, N., Brakken-Thal, C., O'Connor, M.B., King-Jones, K., 2016. The Insect Prothoracic Gland as a Model for Steroid Hormone Biosynthesis and Regulation. Cell Rep 16, 247–262. https://doi.org/10.1016/j.celrep.2016.05.053

Pagès, H., Aboyoun, P., Gentleman, R., 2017. Biostrings: Efficient manipulation of biological strings. R package version 2.46.0.

Pallares, L.F., Harr, B., Turner, L.M., Tautz, D., 2014. Use of a natural hybrid zone for genomewide association mapping of craniofacial traits in the house mouse. Mol Ecol 23, 5756–5770. https://doi.org/10.1111/mec.12968

Palliyil, S., Zhu, J., Baker, L.R., Neuman, S.D., Bashirullah, A., Kumar, J.P., 2018. Allocation of distinct organ fates from a precursor field requires a shift in expression and function of gene regulatory networks. PLoS Genet. 14, e1007185. https://doi.org/10.1371/journal.pgen.1007185

Pan, D., Rubin, G.M., 1995. cAMP-dependent protein kinase and hedgehog act antagonistically in regulating decapentaplegic transcription in drosophila imaginal discs. Cell 80, 543–552. https://doi.org/10.1016/0092-8674(95)90508-1

Panfilio, K.A., 2008. Extraembryonic development in insects and the acrobatics of blastokinesis. Dev. Biol. 313, 471–491. https://doi.org/10.1016/j.ydbio.2007.11.004

Pardue, M.L., Gall, J.G., 1969. MOLECULAR HYBRIDIZATION OF RADIOACTIVE DNA TO THE DNA OF CYTOLOGICAL PREPARATIONS. Proceedings of the National Academy of Sciences 64, 600–604. https://doi.org/10.1073/pnas.64.2.600

Paris, M., Kaplan, T., Li, X.Y., Villalta, J.E., Lott, S.E., Eisen, M.B., 2013. Extensive Divergence of Transcription Factor Binding in Drosophila Embryos with Highly Conserved Gene Expression. PLOS Genetics 9, e1003748. https://doi.org/10.1371/journal.pgen.1003748

Pereira, P.S., Pinho, S., Johnson, K., Couso, J.P., Casares, F., 2006. A 3′cis-regulatory region controlswingless expression in theDrosophila eye and leg primordia. Dev. Dyn. 235, 225–234. https://doi.org/10.1002/dvdy.20606

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J., Quackenbush, J., 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19, 651–652. https://doi.org/10.1093/bioinformatics/btg034

Pesch, Y.-Y., Riedel, D., Patil, K.R., Loch, G., Behr, M., 2016. Chitinases and Imaginal disc growth factors organize the extracellular matrix formation at barrier tissues in insects. Sci Rep 6, 18340. https://doi.org/10.1038/srep18340

Petryk, A., Warren, J.T., Marqués, G., Jarcho, M.P., Gilbert, L.I., Kahler, J., Parvy, J.-P., Li, Y., Dauphin-Villemant, C., O'Connor, M.B., 2003. Shade is the Drosophila P450 enzyme that mediates the hydroxylation of ecdysone to the steroid insect molting hormone 20-hydroxyecdysone. Proc. Natl. Acad. Sci. U.S.A. 100, 13773–13778. https://doi.org/10.1073/pnas.2336088100

Ponting, C.P., Oliver, P.L., Reik, W., 2009. Evolution and Functions of Long Noncoding RNAs. Cell 136, 629–641. https://doi.org/10.1016/j.cell.2009.02.006

Pool, J.E., Aquadro, C.F., 2007. The genetic basis of adaptive pigmentation variation in Drosophila melanogaster: ADAPTIVE PIGMENTATION IN D. MELANOGASTER. Molecular Ecology 16, 2844–2851. https://doi.org/10.1111/j.1365-294X.2007.03324.x

Posnien, N., Hopfen, C., Hilbrant, M., Ramos-Womack, M., Murat, S., Schönauer, A., Herbert, S.L., Nunes, M.D.S., Arif, S., Breuker, C.J., Schlötterer, C., Mitteroecker, P., McGregor, A.P., 2012. Evolution of Eye Morphology and Rhodopsin Expression in the Drosophila melanogaster Species Subgroup. PLoS ONE 7, e37346. https://doi.org/10.1371/journal.pone.0037346

Potier, D., Davie, K., Hulselmans, G., Naval Sanchez, M., Haagen, L., Huynh-Thu, V.A., Koldere, D., Celik, A., Geurts, P., Christiaens, V., Aerts, S., 2014. Mapping Gene Regulatory Networks in Drosophila Eye Development by Large-Scale Transcriptome Perturbations and Motif Inference. Cell Reports 9, 2290–2303. https://doi.org/10.1016/j.celrep.2014.11.038

Prpic, N.-M., 2019. A lesson in homology. eLife 8, e48335. https://doi.org/10.7554/eLife.48335

Prpic, N.-M., Wigand, B., Damen, W., Klingler, M., 2001. Expression of dachshund in wild-type and Distal-less mutant Tribolium corroborates serial homologies in insect appendages. Development Genes and Evolution 211, 467–477. https://doi.org/10.1007/s004270100178

Prud'homme, B., Gompel, N., Carroll, S.B., 2007. Emerging principles of regulatory evolution. Proceedings of the National Academy of Sciences 104, 8605–8612. https://doi.org/10.1073/pnas.0700488104

Prud'homme, B., Gompel, N., Rokas, A., Kassner, V.A., Williams, T.M., Yeh, S.-D., True, J.R., Carroll, S.B., 2006. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. Nature 440, 1050–1053. https://doi.org/10.1038/nature04597

Qu, M., Ma, L., Chen, P., Yang, Q., 2014. Proteomic analysis of insect molting fluid with a focus on enzymes involved in chitin degradation. J. Proteome Res. 13, 2931–2940. https://doi.org/10.1021/pr5000957

Quinlan, A.R., 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics 47, 11.12.1-34. https://doi.org/10.1002/0471250953.bi1112s47

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Quiring, R., Walldorf, U., Kloter, U., Gehring, W., 1994. Homology of the eyeless gene of Drosophila to the Small eye gene in mice and Aniridia in humans. Science 265, 785–789. https://doi.org/10.1126/science.7914031

R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raff, R.A., 2000. Evo-devo: the evolution of a new discipline. Nat Rev Genet 1, 74–79. https://doi.org/10.1038/35049594

Ramaekers, A., Weinberger, S., Claeys, A., Kapun, M., Yan, J., Wolf, R., Flatt, T., Buchner, E., Hassan, B.A., 2018. Altering the temporal regulation of one transcription factor drives sensory trade-offs (preprint). Developmental Biology. https://doi.org/10.1101/348375

Ramain, P., Heitzler, P., Haenlin, M., Simpson, P., 1993. pannier, a negative regulator of achaete and scute in Drosophila, encodes a zinc finger protein with homology to the vertebrate transcription factor GATA-1. Development 119, 1277–1291.

Rathke, H., 1844. Zur Entwicklungsgeschichte der Maulwurfsgrille (Gryllotalpa vulgaris). Arch Anat Physiol wiss Med 27–37.

Rau, A., Gallopin, M., Celeux, G., Jaffrézic, F., 2013. Data-based filtering for replicated high-throughput transcriptome sequencing experiments. Bioinformatics 29, 2146–2152. https://doi.org/10.1093/bioinformatics/btt350

Rau, A., Maugis-Rabusseau, C., 2017. Transformation and model choice for RNA-seq co-expression analysis. Briefings in Bioinformatics bbw128. https://doi.org/10.1093/bib/bbw128

Rebeiz, M., Pool, J.E., Kassner, V.A., Aquadro, C.F., Carroll, S.B., 2009. Stepwise Modification of a Modular Enhancer Underlies Adaptation in a Drosophila Population. Science 326, 1663–1667. https://doi.org/10.1126/science.1178357

Reeder, T.W., Townsend, T.M., Mulcahy, D.G., Noonan, B.P., Wood, P.L., Sites, J.W., Wiens, J.J., 2015. Integrated Analyses Resolve Conflicts over Squamate Reptile Phylogeny and Reveal Unexpected Placements for Fossil Taxa. PLoS ONE 10, e0118199. https://doi.org/10.1371/journal.pone.0118199

Rendeiro, A.F., Schmidl, C., Strefford, J.C., Walewska, R., Davis, Z., Farlik, M., Oscier, D., Bock, C., 2016. Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. Nature Communications 7, 11938. https://doi.org/10.1038/ncomms11938

Rennie, S., Dalby, M., Lloret-Llinares, M., Bakoulis, S., Dalager Vaagensø, C., Heick Jensen, T., Andersson, R., 2018a. Transcription start site analysis reveals widespread divergent transcription in D. melanogaster and core promoter-encoded enhancer activities. Nucleic Acids Research 46, 5455–5469. https://doi.org/10.1093/nar/gky244

Rennie, S., Dalby, M., van Duin, L., Andersson, R., 2018b. Transcriptional decomposition reveals active chromatin architectures and cell specific regulatory interactions. Nat Commun 9, 487. https://doi.org/10.1038/s41467-017-02798-1

Reuveni, E., Getselter, D., Oron, O., Elliott, E., 2018. Differential contribution of cis and trans gene transcription regulatory mechanisms in amygdala and prefrontal cortex and modulation by social stress. Scientific Reports 8. https://doi.org/10.1038/s41598-018-24544-3

Reynolds, S.E., Samuels, R.I., 1996. Physiology and Biochemistry of Insect Moulting Fluid, in: Evans, P.D. (Ed.), Advances in Insect Physiology. Academic Press, pp. 157–232. https://doi.org/10.1016/S0065-2806(08)60031-4

Rifkin, S.A., Kim, J., White, K.P., 2003. Evolution of gene expression in the Drosophila melanogaster subgroup. Nature Genetics 33, 138–144. https://doi.org/10.1038/ng1086

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., Thiessen, N., Griffith, O.L., He, A., Marra, M., Snyder, M., Jones, S., 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nature Methods 4, 651–657. https://doi.org/10.1038/nmeth1068

Rockman, M.V., Kruglyak, L., 2006. Genetics of global gene expression. Nat Rev Genet 7, 862–872. https://doi.org/10.1038/nrg1964

Rogers, W.A., Salomone, J.R., Tacy, D.J., Camino, E.M., Davis, K.A., Rebeiz, M., Williams, T.M., 2013. Recurrent Modification of a Conserved Cis-Regulatory Element Underlies Fruit Fly Pigmentation Diversity. PLOS Genetics 9, e1003740. https://doi.org/10.1371/journal.pgen.1003740

Rong, S., Li, D.-Q., Zhang, X.-Y., Li, S., Zhu, K.Y., Guo, Y.-P., Ma, E.-B., Zhang, J.-Z., 2013. RNA interference to reveal roles of β-N-acetylglucosaminidase gene during molting process in Locusta migratoria. Insect Sci. 20, 109–119. https://doi.org/10.1111/j.1744-7917.2012.01573.x

Roonwal Mithan Lal, Imms Augustus Daniel, 1936. X - Studies on the embryology of the African migratory locust, Locusta migratoria migratorioides R. and F. - I - The early development, with a new theory of multi-phased gastrulation among insects. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 226, 391–421. https://doi.org/10.1098/rstb.1936.0011

Rost, M.M., Poprawa, I., Klag, J., 2004. Ultrastructure of the Pleuropodium in 8-d-old Embryos of Thermobia domestica (Packard) (Insecta, Zygentoma). Ann Entomol Soc Am 97, 541–547. https://doi.org/10.1603/0013-8746(2004)097[0541:UOTPID]2.0.CO;2

S. MILLAM STANLEY, M., W. GRUNDMANN, A., 1970. The Embryonic Development of Tribolium confusum. https://doi.org/10.1093/aesa/63.5.1248

Sarropoulos, I., Marin, R., Cardoso-Moreira, M., Kaessmann, H., 2019. Developmental dynamics of lncRNAs across mammalian organs and species. Nature. https://doi.org/10.1038/s41586-019-1341-x

Sato, M., Saigo, K., 2000. Involvement of pannier and u-shaped in regulation of Decapentaplegic-dependent wingless expression in developing Drosophila notum. Mechanisms of Development 93, 127–138. https://doi.org/10.1016/S0925-4773(00)00282-3

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. Nature Methods 9, 676–682. https://doi.org/10.1038/nmeth.2019

Schluter, D., 2000. The Ecology of Adaptive Radiation, Oxford Series in Ecology and Evolution. Oxford University Press, Oxford, New York.

Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., Talianidis, I., Flicek, P., Odom, D.T., 2010. Five-Vertebrate ChIP-seq Reveals the Evolutionary Dynamics of Transcription Factor Binding 328, 6.

Schmidt-Ott, U., 2000. The amnioserosa is an apomorphic character of cyclorrhaphan flies. Dev. Genes Evol. 210, 373–376. https://doi.org/10.1007/s004270050325

Schulz, M.H., Zerbino, D.R., Vingron, M., Birney, E., 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28, 1086–1092. https://doi.org/10.1093/bioinformatics/bts094

Schwager, E.E., Sharma, P.P., Clarke, T., Leite, D.J., Wierschin, T., Pechmann, M., Akiyama-Oda, Y., Esposito, L., Bechsgaard, J., Bilde, T., Buffry, A.D., Chao, H., Dinh, H., Doddapaneni, H., Dugan, S., Eibner, C., Extavour, C.G., Funch, P., Garb, J., Gonzalez, L.B., Gonzalez, V.L., Griffiths-Jones, S., Han, Y., Hayashi, C., Hilbrant, M., Hughes, D.S.T., Janssen, R., Lee, S.L., Maeso, I., Murali, S.C., Muzny, D.M., Nunes da Fonseca, R., Paese, C.L.B., Qu, J., Ronshaugen, M., Schomburg, C., Schönauer, A., Stollewerk, A., Torres-Oliva, M., Turetzek, N., Vanthournout, B., Werren, J.H., Wolff, C., Worley, K.C., Bucher, G., Gibbs, R.A., Coddington, J., Oda, H., Stanke, M., Ayoub, N.A.,

Prpic, N.-M., Flot, J.-F., Posnien, N., Richards, S., McGregor, A.P., 2017. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. BMC Biology 15. https://doi.org/10.1186/s12915-017-0399-x

Scott, M.P., Weiner, A.J., 1984. Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. Proceedings of the National Academy of Sciences 81, 4115–4119. https://doi.org/10.1073/pnas.81.13.4115

Seeds, A.M., Tsui, M.M., Sunu, C., Spana, E.P., York, J.D., 2015. Inositol phosphate kinase 2 is required for imaginal disc development in *Drosophila*. Proc Natl Acad Sci USA 201514684. https://doi.org/10.1073/pnas.1514684112

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., Cavalli, G., 2012. Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. Cell 148, 458–472. https://doi.org/10.1016/j.cell.2012.01.010

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504. https://doi.org/10.1101/gr.1239303

Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jónsson, B., Schluter, D., Kingsley, D.M., 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature 428, 717–723. https://doi.org/10.1038/nature02415

Shi, Y., Noll, M., 2009. Determination of cell fates in the R7 equivalence group of the Drosophila eye by the concerted regulation of D-Pax2 and TTK88. Developmental Biology 331, 68–77. https://doi.org/10.1016/j.ydbio.2009.04.026

Shutts, J.H., 1952. Some characteristics of the hatching enzyme in the eggs of Melanoplus differenitalis (Thomas). Proc S Dak Acad Sci. 31, 158–163.

Siddall, N.A., Hime, G.R., Pollock, J.A., Batterham, P., 2009. Ttk69-dependent repression of lozenge prevents the ectopic development of R7 cells in the Drosophila larval eye disc. BMC Dev Biol 9, 64. https://doi.org/10.1186/1471-213X-9-64

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Singh, A., Chan, J., Chern, J.J., Choi, K.-W., 2005. Genetic interaction of Lobe with its modifiers in dorsoventral patterning and growth of the Drosophila eye. Genetics 171, 169–183. https://doi.org/10.1534/genetics.105.044180

Singh, A., Choi, K.-W., 2003. Initial state of the Drosophila eye before dorsoventral specification is equivalent to ventral. Development 130, 6351–6360. https://doi.org/10.1242/dev.00864

Singh, P.P., Arora, J., Isambert, H., 2015. Identification of Ohnolog Genes Originating from Whole Genome Duplication in Early Vertebrates, Based on Synteny Comparison across Multiple Genomes. PLOS Computational Biology 11, e1004394. https://doi.org/10.1371/journal.pcbi.1004394

Siomava, N., Wimmer, E.A., Posnien, N., 2016. Size relationships of different body parts in the three dipteran species Drosophila melanogaster, Ceratitis capitata and Musca domestica. Development Genes and Evolution 226, 245–256. https://doi.org/10.1007/s00427-016-0543-6

Slifer, E.H., 1938. A cytological study of the pleuropodia of Melanoplus differentialis (Orthoptera, Acrididae) which furnishes new evidence that they produce the hatching enzyme. J. Morphol. 63, 181–205. https://doi.org/10.1002/jmor.1050630109

Snodgrass, R.E., 1938. Evolution of the annelida onychophora and arthropoda.

Sonawane, A.R., Platig, J., Fagny, M., Chen, C.-Y., Paulson, J.N., Lopes-Ramos, C.M., DeMeo, D.L., Quackenbush, J., Glass, K., Kuijjer, M.L., 2017. Understanding Tissue-Specific Gene Regulation. Cell Rep 21, 1077–1088. https://doi.org/10.1016/j.celrep.2017.10.001

Sorrentino, R.P., Tokusumi, T., Schulz, R.A., 2007. The Friend of GATA protein U-shaped functions as a hematopoietic tumor suppressor in Drosophila. Developmental Biology 311, 311–323. https://doi.org/10.1016/j.ydbio.2007.08.011

Stanojevic, D., Small, S., Levine, M., 1991. Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. Science 254, 1385–1387. https://doi.org/10.1126/science.1683715

Starks, R.R., Biswas, A., Jain, A., Tuteja, G., 2019. Combined analysis of dissimilar promoter accessibility and gene expression profiles identifies tissue-specific genes and actively repressed networks. Epigenetics & Chromatin 12, 16. https://doi.org/10.1186/s13072-019-0260-2

Stay, B., 1977. Fine structure of two types of pleuropodia in Diploptera punctata (Dictyoptera: Blaberidae) with observations on their permeability. International Journal of Insect Morphology and Embryology 6, 67–95. https://doi.org/10.1016/0020-7322(77)90013-7

Stern, D.L., 2014. Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. Trends Genet. 30, 547–554. https://doi.org/10.1016/j.tig.2014.09.006

Stern, D.L., Orgogozo, V., 2008. THE LOCI OF EVOLUTION: HOW PREDICTABLE IS GENETIC EVOLUTION? Evolution 62, 2155–2177. https://doi.org/10.1111/j.1558-5646.2008.00450.x

Struhl, G., Basler, K., 1993. Organizing activity of wingless protein in Drosophila. Cell 72, 527–540. https://doi.org/10.1016/0092-8674(93)90072-x

Strutt, D.I., Wiersdorff, V., Mlodzik, M., 1995. Regulation of furrow progression in the Drosophila eye by cAMP-dependent protein kinase A. Nature 373, 705–709. https://doi.org/10.1038/373705a0

Sucena, E., Delon, I., Jones, I., Payre, F., Stern, D.L., 2003. Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. Nature 424, 935–938. https://doi.org/10.1038/nature01768

Sucena, É., Stern, D.L., 2000. Divergence of larval morphology between Drosophila sechellia and its sibling species caused by cis-regulatory evolution of ovo/shaven-baby. PNAS 97, 4530–4534. https://doi.org/10.1073/pnas.97.9.4530

Sugahara, R., Tanaka, S., Shiotsuki, T., 2017. RNAi-mediated knockdown of SPOOK reduces ecdysteroid titers and causes precocious metamorphosis in the desert locust Schistocerca gregaria. Dev. Biol. 429, 71–80. https://doi.org/10.1016/j.ydbio.2017.07.007

Sui, Y.-P., Liu, X.-B., Chai, L.-Q., Wang, J.-X., Zhao, X.-F., 2009. Characterization and influences of classical insect hormones on the expression profiles of a molting carboxypeptidase A from the cotton bollworm (Helicoverpa armigera). Insect Mol. Biol. 18, 353–363. https://doi.org/10.1111/j.1365-2583.2009.00879.x

Suster, M.L., Seugnet, L., Bate, M., Sokolowski, M.B., 2004. Refining GAL4-driven transgene expression in Drosophila with a GAL80 enhancer-trap. genesis 39, 240–245. https://doi.org/10.1002/gene.20051

Suvorov, A., Nolte, V., Pandey, R.V., Franssen, S.U., Futschik, A., Schlötterer, C., 2013. Intra-Specific Regulatory Variation in Drosophila pseudoobscura. PLoS ONE 8, e83547. https://doi.org/10.1371/journal.pone.0083547

Swalla, B.J., Smith, A.B., 2008. Deciphering deuterostome phylogeny: molecular, morphological and palaeontological perspectives. Philosophical Transactions of the Royal Society B: Biological Sciences 363, 1557–1568. https://doi.org/10.1098/rstb.2007.2246

Tan, S., Amos, W., Laughlin, S.B., 2005. Captivity selects for smaller eyes. Current Biology 15, R540–R542. https://doi.org/10.1016/j.cub.2005.07.019

Tanaka, M., Kobayashi, Y.K., Ando, H., 1985. Embryonic Development of the Nervous System and Other Ectodermal Derivatives in the Primitive Moth, Endoclita sinensis (Lepidoptera, Hepialidae) 15.

Tanizawa, T., Ando, H., Tojo, K., 2007. Notes on the Pleuropodia in the Giant Water Bug Appasus japonicus (Heteroptera, Belostomatida). Proc. Arthropod. Embryol. Soc. Jpn. 42, 9–11.

Tarazona, O.A., Lopez, D.H., Slota, L.A., Cohn, M.J., 2019. Evolution of limb development in cephalopod mollusks. eLife 8, e43828. https://doi.org/10.7554/eLife.43828

Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., Conesa, A., 2011. Differential expression in RNA-seq: a matter of depth. Genome Res. 21, 2213–2223. https://doi.org/10.1101/gr.124321.111

Tautz, D., 2000. Evolution of transcriptional regulation. Curr. Opin. Genet. Dev. 10, 575–579.

Tear, G., Akam, M., Martinez-Arias, A., 1990. Isolation of an abdominal-A gene from the locust Schistocerca gregaria and its expression during early embryogenesis. Development 110, 915–925.

Technau, G.M., Berger, C., Urbach, R., 2006. Generation of cell diversity and segmental pattern in the embryonic central nervous system of Drosophila. Dev. Dyn. 235, 861–869. https://doi.org/10.1002/dvdy.20566

The Gene Ontology Consortium, 2019. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Research 47, D330–D338. https://doi.org/10.1093/nar/gky1055

Thompson, D., Regev, A., Roy, S., 2015. Comparative Analysis of Gene Regulatory Networks: From Network Reconstruction to Evolution. Annual Review of Cell and Developmental Biology 31, 399–428. https://doi.org/10.1146/annurev-cellbio-100913-012908

Todd, E.V., Black, M.A., Gemmell, N.J., 2016. The power and promise of RNA-seq in ecology and evolution. Mol Ecol 25, 1224–1241. https://doi.org/10.1111/mec.13526

Tomoyasu, Y., Ueno, N., Nakamura, M., 2000. The Decapentaplegic morphogen gradient regulates the notal wingless expression through induction of pannier and u-shaped in Drosophila. Mechanisms of Development 96, 37–49. https://doi.org/10.1016/S0925-4773(00)00374-9

Torres-Oliva, M., 2016. Identification of the molecular changes underlying head morphology variation in closely related Drosophila species. Georg-August-Universität Göttingen, Göttingen.

Torres-Oliva, M., Almudi, I., McGregor, A.P., Posnien, N., 2016. A robust (re-)annotation approach to generate unbiased mapping references for RNA-seq-based analyses of differential expression across closely related species. BMC Genomics 17. https://doi.org/10.1186/s12864-016-2646-x

Torres-Oliva, M., Schneider, J., Wiegleb, G., Kaufholz, F., Posnien, N., 2018. Dynamic genome wide expression profiling of Drosophila head development reveals a novel role of Hunchback in retinal glia cell development and blood-brain barrier integrity. PLOS Genetics 14, e1007180. https://doi.org/10.1371/journal.pgen.1007180

Treisman, J.E., 2013. Retinal differentiation in *Drosophila*. WIREs Dev Biol 2, 545–557. https://doi.org/10.1002/wdev.100

Treisman, J.E., Heberlein, U., 1998. Eye development in Drosophila: formation of the eye field and control of differentiation. Curr. Top. Dev. Biol. 39, 119–158.

Treisman, J.E., Rubin, G.M., 1995. wingless inhibits morphogenetic furrow movement in the Drosophila eye disc. Development 121, 3519–3527.

Tsui, M.M., York, J.D., 2010. Roles of inositol phosphates and inositol pyrophosphates in development, cell signaling and nuclear processes. Advances in Enzyme Regulation 50, 324–337. https://doi.org/10.1016/j.advenzreg.2009.12.002

Tsutsumi, T., 2008. Embryonic development of a snakefly, Inocellia japonica Okamoto. Proc. Arthropod Embryol Soc Jpn.

Turetzek, N., Khadjeh, S., Schomburg, C., Prpic, N.-M., 2017. Rapid diversification of homothorax expression patterns after gene duplication in spiders. BMC Evolutionary Biology 17. https://doi.org/10.1186/s12862-017-1013-0

Turetzek, N., Pechmann, M., Schomburg, C., Schneider, J., Prpic, N.-M., 2016. Neofunctionalization of a Duplicate *dachshund* Gene Underlies the Evolution of a Novel Leg Segment in Arachnids. Molecular Biology and Evolution 33, 109–121. https://doi.org/10.1093/molbev/msv200

Turner, B.M., Birley, A.J., Lavender, J., 1992. Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei. Cell 69, 375–384. https://doi.org/10.1016/0092-8674(92)90417-b

Uchifune, T., Machida, R., 2005. Embryonic development of Galloisiana yuasai Asahina, with special reference to external morphology (insecta: Grylloblattodea). J. Morphol. 266, 182–207. https://doi.org/10.1002/jmor.10373

UniProt: a worldwide hub of protein knowledge, 2019. . Nucleic Acids Res 47, D506–D515. https://doi.org/10.1093/nar/gky1049

Uyehara, C.M., Nystrom, S.L., Niederhuber, M.J., Leatham-Jensen, M., Ma, Y., Buttitta, L.A., McKay, D.J., 2017. Hormone-dependent control of developmental timing through regulation of chromatin accessibility. Genes Dev. 31, 862–875. https://doi.org/10.1101/gad.298182.117

van Berkum, N.L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L.A., Dekker, J., Lander, E.S., 2010. Hi-C: A Method to Study the Three-dimensional Architecture of Genomes. Journal of Visualized Experiments. https://doi.org/10.3791/1869

Viscuso, R., Sottile, L., 2008. Fine structure of pleuropodia in three species of Insecta Orthoptera during embryonic development. Italian Journal of Zoology 75, 11–19. https://doi.org/10.1080/11250000701690475

Wang, C.-W., Sun, Y.H., 2012. Segregation of eye and antenna fates maintained by mutual antagonism in Drosophila. Development 139, 3413–3421. https://doi.org/10.1242/dev.078857

Wang, Y.-H., Huang, M.-L., 2009. Reduction of Lobe leads to TORC1 hypoactivation that induces ectopic Jak/STAT signaling to impair Drosophila eye development. Mechanisms of Development 126, 781–790. https://doi.org/10.1016/j.mod.2009.08.005

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10, 57–63. https://doi.org/10.1038/nrg2484

Warren, J.T., Petryk, A., Marques, G., Jarcho, M., Parvy, J.-P., Dauphin-Villemant, C., O'Connor, M.B., Gilbert, L.I., 2002. Molecular and biochemical characterization of two P450 enzymes in the ecdysteroidogenic pathway of Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.A. 99, 11043–11048. https://doi.org/10.1073/pnas.162375799

Warren, J.T., Petryk, A., Marqués, G., Parvy, J.-P., Shinoda, T., Itoyama, K., Kobayashi, J., Jarcho, M., Li, Y., O'Connor, M.B., Dauphin-Villemant, C., Gilbert, L.I., 2004. Phantom encodes the 25-hydroxylase of Drosophila melanogaster and Bombyx mori: a P450 enzyme critical in ecdysone biosynthesis. Insect Biochem. Mol. Biol. 34, 991–1010. https://doi.org/10.1016/j.ibmb.2004.06.009

Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., Zdobnov, E.M., 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msx319

Weber, U., Paricio, N., Mlodzik, M., 2000. Jun mediates Frizzled-induced R3/R4 cell fate distinction and planar polarity determination in the Drosophila eye. Development 127, 3619–3629.

Wei, Z., Yin, Y., Zhang, B., Wang, Z., Peng, G., Cao, Y., Xia, Y., 2007. Cloning of a novel protease required for the molting of Locusta migratoria manilensis. Development, Growth & Differentiation 49, 611–621. https://doi.org/10.1111/j.1440-169X.2007.00957.x

Wheeler, W.M., 1890. On the appendages of the first abdominal segment of embryo insects. s.n.], [Madison.

Winick, J., Abel, T., Leonard, M.W., Michelson, A.M., Chardon-Loriaux, I., Holmgren, R.A., Maniatis, T., Engel, J.D., 1993. A GATA family transcription factor is expressed along the embryonic dorsoventral axis in Drosophila melanogaster. Development 119, 1055–1065.

Wittkopp, P.J., 2013. V.7. Evolution of Gene Expression, in: Losos, J.B., Baum, D.A., Futuyma, D.J., Hoekstra, H.E., Lenski, R.E., Moore, A.J., Peichel, C.L., Schluter, D., Whitlock, M.C. (Eds.), The Princeton Guide to Evolution. Princeton University Press, Princeton. https://doi.org/10.1515/9781400848065-058

Wittkopp, P.J., 2005. Genomic sources of regulatory variation in cis and in trans. Cellular and Molecular Life Sciences 62, 1779–1783. https://doi.org/10.1007/s00018-005-5064-9

Wittkopp, P.J., Haerum, B.K., Clark, A.G., 2008. Regulatory changes underlying expression differences within and between Drosophila species. Nat Genet 40, 346–350. https://doi.org/10.1038/ng.77

Wittkopp, P.J., Haerum, B.K., Clark, A.G., 2004. Evolutionary changes in cis and trans gene regulation. Nature 430, 85–88. https://doi.org/10.1038/nature02698

Wolpert, L., Tickle, C., 2011. Principles of development, 4th ed. / L. Wolpert, Cheryll Tickle. ed. Oxford : Oxford University Press.

Wray, G.A., 2007. The evolutionary significance of *cis*-regulatory mutations. Nature Reviews Genetics 8, 206–216. https://doi.org/10.1038/nrg2063

Wray, G.A., 2003. The Evolution of Transcriptional Regulation in Eukaryotes. Molecular Biology and Evolution 20, 1377–1419. https://doi.org/10.1093/molbev/msg140

Xi, Y., Pan, P.-L., Ye, Y.-X., Yu, B., Xu, H.-J., Zhang, C.-X., 2015. Chitinase-like gene family in the brown planthopper, Nilaparvata lugens. Insect Mol. Biol. 24, 29–40. https://doi.org/10.1111/imb.12133

Xie, K.T., Wang, G., Thompson, A.C., Wucherpfennig, J.I., Reimchen, T.E., MacColl, A.D.C., Schluter, D., Bell, M.A., Vasquez, K.M., Kingsley, D.M., 2019. DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. Science 363, 81–84. https://doi.org/10.1126/science.aan1425

Xiong, W.C., Montell, C., 1993. tramtrack is a transcriptional repressor required for cell fate determination in the Drosophila eye. Genes Dev. 7, 1085–1096. https://doi.org/10.1101/gad.7.6.1085

Yamamoto, D.S., Sumitani, M., Tojo, K., Lee, J.M., Hatakeyama, M., 2004. Cloning of a decapentaplegic orthologue from the sawfly, Athalia rosae (Hymenoptera), and its expression in the embryonic appendages. Dev. Genes Evol. 214, 128–133. https://doi.org/10.1007/s00427-004-0387-3

Yu, H., Luscombe, N.M., Qian, J., Gerstein, M., 2003. Genomic analysis of gene expression relationships in transcriptional regulatory networks. Trends in Genetics 19, 422–427. https://doi.org/10.1016/S0168-9525(03)00175-6

Yu, J., Pacifico, S., Liu, G., Finley, R.L., 2008. DroID: the Drosophila Interactions Database, a comprehensive resource for annotated gene and protein interactions. BMC Genomics 9, 461. https://doi.org/10.1186/1471-2164-9-461

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829. https://doi.org/10.1101/gr.074492.107

Zhang, H., Shinmyo, Y., Mito, T., Miyawaki, K., Sarashina, I., Ohuchi, H., Noji, S., 2005. Expression patterns of the homeotic genes Scr, Antp, Ubx, and abd-A during embryogenesis of the cricket Gryllus bimaculatus. Gene Expr. Patterns 5, 491–502. https://doi.org/10.1016/j.modgep.2004.12.006

Zhang, J., Lu, A., Kong, L., Zhang, Q., Ling, E., 2014. Functional analysis of insect molting fluid proteins on the protection and regulation of ecdysis. J. Biol. Chem. 289, 35891–35906. https://doi.org/10.1074/jbc.M114.599597

Zhang, X., Borevitz, J.O., 2009. Global Analysis of Allele-Specific Expression in *Arabidopsis thaliana*. Genetics 182, 943–954. https://doi.org/10.1534/genetics.109.103499

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biol 9, R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zheng, L., Carthew, R.W., 2008. Lola regulates cell fate by antagonizing Notch induction in the Drosophila eye. Mechanisms of Development 125, 18–29. https://doi.org/10.1016/j.mod.2007.10.007

Zheng, W., Zhao, H., Mancera, E., Steinmetz, L.M., Snyder, M., 2010. Genetic analysis of variation in transcription factor binding in yeast. Nature 464, 1187–1191. https://doi.org/10.1038/nature08934

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., Chanda, S.K., 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature Communications 10, 1523. https://doi.org/10.1038/s41467-019-09234-6

Zhu, J., Sanborn, J.Z., Diekhans, M., Lowe, C.B., Pringle, T.H., Haussler, D., 2007. Comparative Genomics Search for Losses of Long-Established Genes on the Human Lineage. PLoS Computational Biology 3, 12.

Zhu, K.Y., Merzendorfer, H., Zhang, W., Zhang, J., Muthukrishnan, S., 2016. Biosynthesis, Turnover, and Functions of Chitin in Insects. Annu. Rev. Entomol. 61, 177–196. https://doi.org/10.1146/annurev-ento-010715-023933

Zhu, Q., Arakane, Y., Beeman, R.W., Kramer, K.J., Muthukrishnan, S., 2008. Functional specialization among insect chitinase family genes revealed by RNA interference. Proc. Natl. Acad. Sci. U.S.A. 105, 6650–6655. https://doi.org/10.1073/pnas.0800739105